

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 20:09:21 ; Search time 14 Seconds
(without alignments)
2882.455 Million cell updates/sec

Title: US-10-049-745-4

Perfect score: 775

Sequence: 1 MTIVDKASSSDPAYQNP.....AAESLEEDAAASLPFFSEG 775

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 10

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	565	72.9	1198	1 UB42_HUMAN	Q9h9j4 homo sapien
2	14	1.8	1121	1 UB36_HUMAN	Q9p275 homo sapien

ALIGNMENTS

RESULT 1
UB42_HUMAN
ID UB42_HUMAN STANDARD; PRT; 1198 AA.
AC Q9H9J4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Ubiquitin carboxyl-terminal hydrolase 42 (EC 3.1.2.15) (Ubiquitin
thiolesterase 42) (Ubiquitin-specific processing protease 42)
GN (Deubiquitinating enzyme 42) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayaashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
RA Niimiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -1- SIMILARITY: Belongs to peptidase family C19.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AK022759; BAB14232.1; -.
CC MEROPS; C19.048; -.
CC Genew; HGNC:20068; USP42.
CC InterPro; IPR001394; Peptidase_C19.
CC Pfam; PF00443; UCH; 1.
CC PROSITE; PS00972; UCH_2_1; FALSE_NEG.
CC PROSITE; PS00973; UCH_2_2; 1.
CC PROSITE; PS0235; UCH_2_3; 1.
CC KW Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
CC FT ACT_SITE 120 120 BY SIMILARITY.
CC FT ACT_SITE 362 362 BY SIMILARITY.
CC FT ACT_SITE 371 371 BY SIMILARITY.
CC FT NON_TER 1198 1198
CC SQ SEQUENCE 1198 AA; 130587 MW; CF0FDB5184AE9536 CRC64;
Query Match 72.9%; Score 565; DB 1; Length 1198;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 765; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MTIVDKASSSDPAYQNPQGSSEAVSPGMDAGSASWGAVSLNDVSNHTLSLGPVPGA 60
DB 1 MTIVDKASSSDPAYQNPQGSSEAVSPGMDAGSASWGAVSLNDVSNHTLSLGPVPGA 60
QY 61 VYSSSSVDPKSKPSQKQDQALGDGIAPQKYLFPSEKICLKWQTHRVGAGLQNLGNTC 120
DB 61 VYSSSSVDPKSKPSQKQDQALGDGIAPQKYLFPSEKICLKWQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLITTPPLANYMLSHHSKTCCHAEFCWMTQAHITQALSNPQGVIKPMFVI 180
DB 121 FANAALQCLITTPPLANYMLSHHSKTCCHAEFCWMTQAHITQALSNPQGVIKPMFVI 180
QY 181 NEMRIARHLRFGNQEDAHEFLQYTVDMQKACLNGSNKLDRTHTQATTLVCOIFGGYLS 240
DB 181 NEMRIARHLRFGNQEDAHEFLQYTVDMQKACLNGSNKLDRTHTQATTLVCOIFGGYLS 240
QY 241 RVKCLNCKGVSDTDPYLDITILEIKAAQSVNKALQFVKPEQLDGENSYKSKCKMWPA 300
DB 241 RVKCLNCKGVSDTDPYLDITILEIKAAQSVNKALQFVKPEQLDGENSYKSKCKMWPA 300
QY 301 SKRFTIHRSSNVLTSLKRFANFTGGKIAKADVKYPEYLDIRPYMOPNGEPIVYVLYAVL 360
DB 301 SKRFTIHRSSNVLTSLKRFANFTGGKIAKADVKYPEYLDIRPYMOPNGEPIVYVLYAVL 360
QY 361 VHTGFNCHAGHYFCYIKASNGLWYQNDISVSTSDIRSVLSQQAAYVLFYIRSHDVKNGE 420
DB 361 VHTGFNCHAGHYFCYIKASNGLWYQNDISVSTSDIRSVLSQQAAYVLFYIRSHDVKNGE 420
QY 421 LTHPHTSPQSSPRPVISORVVVTKQAAPGFTGPOLPSHMIKNPPLHNGTGLKDTSPSS 480
DB 421 LTHPHTSPQSSPRPVISORVVVTKQAAPGFTGPOLPSHMIKNPPLHNGTGLKDTSPSS 480
QY 481 MSSPNGNSVNRASFPVNASASVQNVSSVIRPEHPKKQKITIISHNKLPVRCQSQPN 540
DB 481 MSSPNGNSVNRASFPVNASASVQNVSSVIRPEHPKKQKITIISHNKLPVRCQSQPN 540
QY 541 LHSNSLENPTKVPSSSTITNSAVQSTNSASTMSVSSKVTKEIPRSESCSQPVMNGSKLN 600
DB 541 LHSNSLENPTKVPSSSTITNSAVQSTNSASTMSVSSKVTKEIPRSESCSQPVMNGSKLN 600
QY 601 SSVLVPGVAESSEDSDESKGLKNGKGTIVSSHSPQDADEDEATPHELOEPMTLNGA 660
DB 601 SSVLVPGVAESSEDSDESKGLKNGKGTIVSSHSPQDADEDEATPHELOEPMTLNGA 660

```

DR InterPro: IPR001394; Peptidase_C19.
DR Pfam: PF00443; UCH; 1.
DR PROSITE: PS00972; UCH_2_1; 1.
DR PROSITE: PS00973; UCH_2_2; 1.
DR PROSITE: PS02035; UCH_2_3; 1.
DR Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family;
KW Nuclear protein.
KW ACT_SITE 131 131 BY SIMILARITY.
KW ACT_SITE 131 373 BY SIMILARITY.
KW ACT_SITE 382 382 BY SIMILARITY.
KW CONFLICT 573 573 D -> G (IN REF. 2).
KW CONFLICT 806 806 R -> Q (IN REF. 3).
KW CONFLICT 828 828 K -> R (IN REF. 3).
KW CONFLICT 958 958 K -> KKK (IN REF. 3).
KW SEQUENCE 1121 AA; 122626 MW; 2476F5128CBAB5A5 CRC64;

SQ
Query Match 1.8%; Score 14; DB 1; Length 1121;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 310 SNVLTSLKRFANF 323
|||||
DB 321 SNVLTSLKRFANF 334
|||||

Search completed: August 10, 2004, 20:15:42
Job time : 15 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:59:54 ; Search time 50 Seconds
(without alignments)
4862.086 Million cell updates/sec

Title: US-10-049-745-4

Perfect score: 4070

Sequence: 1 MTIWDKASESDPSAYQNP.....AAESLEEDPAASLFFPSEG 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:**

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	4028	99.0	1251	12	US-10-399-645-7
2	4028	99.0	1316	15	US-10-379-981-9
3	3399	98.3	762	14	US-10-163-547-1
4	3198.5	78.6	1326	15	US-10-379-981-12
5	1423.5	35.0	1123	12	US-09-972-525-1
6	1150.5	28.3	565	12	US-10-371-905A-20
7	1148	28.2	530	14	US-10-107-695B-2
8	1142.5	28.1	530	12	US-10-371-905A-16
9	1139.5	28.0	530	12	US-10-371-905A-12
10	1139.5	28.0	574	12	US-10-371-905A-14
11	1135.5	27.9	530	12	US-10-371-905A-18
12	1132.5	27.8	715	12	US-10-371-905A-22
13	1128.5	27.7	1021	12	US-10-371-905A-8
14	1126.5	27.7	633	12	US-10-371-905A-24
15	1125	27.6	472	16	US-10-408-765A-2752

16	1121.5	27.6	1016	12	US-10-371-905A-4
17	1118.5	27.5	530	12	US-10-371-905A-10
18	1114	27.4	398	12	US-10-371-905A-42
19	1113.5	27.4	530	10	US-09-372-348-11
20	1111.5	27.3	530	12	US-10-371-905A-6
21	1096	26.9	530	12	US-10-371-905A-34
22	1073	26.4	530	12	US-10-371-905A-26
23	1052	25.8	530	12	US-10-371-905A-32
24	1049.5	25.8	530	12	US-10-371-905A-36
25	1037	25.3	357	12	US-10-371-905A-40
26	1030	25.3	530	12	US-10-371-905A-30
27	980.5	24.1	497	12	US-10-371-905A-28
28	914	22.5	748	10	US-09-372-348-15
29	898	21.8	545	10	US-09-372-348-17
30	846.5	20.8	526	10	US-09-372-348-16
31	799.5	19.6	314	15	US-10-274-639-11
32	799.5	19.6	314	16	US-10-333-574-11
33	787	19.3	844	16	US-10-437-963-132523
34	786.5	19.3	928	12	US-10-425-114-65150
35	760	18.7	457	15	US-10-369-493-22645
36	735.5	18.1	993	16	US-10-437-963-182980
37	728	17.9	549	12	US-10-424-599-233768
38	719	17.7	1003	12	US-10-425-114-47268
39	718	17.6	1204	16	US-10-437-963-161587
40	701	17.2	1010	16	US-10-437-963-131687
41	690	17.0	735	12	US-10-425-114-54385
42	666	16.4	919	16	US-10-437-963-183316
43	658.5	16.2	636	16	US-10-437-963-116674
44	657.5	16.2	563	12	US-10-425-114-48696
45	652.5	16.0	248	12	US-10-371-905A-46

ALIGNMENTS

RESULT 1

US-10-399-645-7
; Sequence 7, Application US/10399645
; Publication No. US20040029249A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; LEE, Ernestine A.
; APPLICANT: HAFALTA, April J.A.; YUE, Henry
; APPLICANT: LAL, Freeti G.; YAO, Monique G.
; APPLICANT: LU, Yan; CHAWLA, Narinder K.
; APPLICANT: WARREN, Bridget A.; LU, Dying Aina M.
; APPLICANT: BAUGHN, Mariah R.; DELEGEANE, Angelo M.
; APPLICANT: BURFORD, Neil; BOROWSKY, Mark L.
; APPLICANT: LEE, Sally; XU, Yuming
; APPLICANT: GRIFFIN, Jennifer A.; KALLICK, Deborah A.
; APPLICANT: GANDHI, Ameena R.; ARVIZU, Chandra S.
; APPLICANT: ISON, Craig H.; TANG, Y. Tom
; APPLICANT: AZIMZAI, Yalda; ELLIOTT, Vicki S.
; APPLICANT: SWARNAKAR, Anita; RAMKUMAR, Jayalaxmi
; APPLICANT: NGUYEN, Dannel B.; TRIBOULEY, Catherine M.
; APPLICANT: LO, Terence P.; AU-YOUNG, Janice K.
; APPLICANT: THANGAVELU, Kavitha; KEARNEY, Liam
; TITLE OF INVENTION: PROTEASES
; FILE REFERENCE: PI-0263 USN
; CURRENT APPLICATION NUMBER: US/10399,645
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: PCT/US01/51034
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: US 60/241,573
; PRIOR FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US 60/243,643
; PRIOR FILING DATE: 2000-10-25
; PRIOR APPLICATION NUMBER: US 60/245,256
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/248,395
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: US 60/249,826
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/252,303

```
; Sequence 9, Application US/10379981
; Publication No. US20040001820A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pharmaceuticals, Inc.
; APPLICANT: Hahn, Chang S
; APPLICANT: Liu, Hong S
; TITLE OF INVENTION: HUMAN DEUBIQUITINATING PROTEASE GENE ON CHROMOSOME 7 AND ITS
; FILE REFERENCE: USAV2002-0022 WO PCT
; CURRENT APPLICATION NUMBER: US/10/379,981
; PRIOR FILING DATE: 2003-03-05
; PRIOR APPLICATION NUMBER: GB 0218518.9
; PRIOR FILING DATE: 2002-09-08
; PRIOR APPLICATION NUMBER: US 60/366,601
; PRIOR FILING DATE: 2002-03-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 9
; LENGTH: 1316
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-379-981-9

Query Match          99.3%; Score 4028; DB 15; Length 1316;
Best Local Similarity 99.7%; Pred. No. 5.9e-286;
Matches 767; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTIVDKASSSDPSAYONQPGSEAVSPGDMAGSAGWAVSSLNDVSNHTLSLGPVPGA 60
DB 1 MTIVDKASSSDPSAYONQPGSEAVSPGDMAGSAGWAVSSLNDVSNHTLSLGPVPGA 60
QY 61 VVYSSSVDPKSPKQDQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
DB 61 VVYSSSVDPKSPKQDQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTTPPLANYMLSHEHSKTCCHAEFCMCTQAHITQALSNPGDVVKPMFVI 180
DB 121 FANAALQCLTYTTPPLANYMLSHEHSKTCCHAEFCMCTQAHITQALSNPGDVVKPMFVI 180
QY 181 NEMRRIARHLRFNGQDAHEFLQYTVDMQKACLNGSNKLDHRTQATTLVCOIFGGYLR 240
DB 181 NEMRRIARHLRFNGQDAHEFLQYTVDMQKACLNGSNKLDHRTQATTLVCOIFGGYLR 240
QY 241 RVKCLNCKGVSDTFDPLVDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSCKCKMWPA 300
DB 241 RVKCLNCKGVSDTFDPLVDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSCKCKMWPA 300
QY 301 SKRFTIHRSSNVLTLSLKRFPANFTGGKIAKDVKYPEYLDIRPYMSQNGEPIVVVLAVL 360
DB 301 SKRFTIHRSSNVLTLSLKRFPANFTGGKIAKDVKYPEYLDIRPYMSQNGEPIVVVLAVL 360
QY 361 VHTGFNCHAGHYFCYIKASNGLWTQMNDISVSTSDIRSVLSQQAAYLVFIYRSHDVKNNGE 420
DB 361 VHTGFNCHAGHYFCYIKASNGLWTQMNDISVSTSDIRSVLSQQAAYLVFIYRSHDVKNNGE 420
QY 421 LTHPTHSPGQSSPRPVISQSVVTKNQAAAGFIGPQLPSHMIKNPPLHNGTGPKLDTPESS 480
DB 421 LTHPTHSPGQSSPRPVISQSVVTKNQAAAGFIGPQLPSHMIKNPPLHNGTGPKLDTPESS 480
QY 481 MSSPNGNSVNRASFPVNASAVQNSVNRSSVIEHPKKQKITIISIHKLPRVQCOSQPN 540
DB 481 MSSPNGNSVNRASFPVNASAVQNSVNRSSVIEHPKKQKITIISIHKLPRVQCOSQPN 540
QY 541 LHSNSLENPTKVPSPSTITNSAVOSTSNASTMSVSSKVTKEIPRSESCSQPMNGKSKLN 600
DB 541 LHSNSLENPTKVPSPSTITNSAVOSTSNASTMSVSSKVTKEIPRSESCSQPMNGKSKLN 600
QY 601 SSVLPVYGAESSEDSDESKGLGKENGIGTIVSSHSPQDQDADEEATPHELQEPMTLNGA 660
DB 601 SSVLPVYGAESSEDSDESKGLGKENGIGTIVSSHSPQDQDADEEATPHELQEPMTLNGA 660
QY 661 NSADSDDPKENGLAPDGASCOGQPALHSENPFPAKANGLPGLMPAPLLSLPEDKILETF 720
DB 661 NSADSDDPKENGLAPDGASCOGQPALHSENPFPAKANGLPGLMPAPLLSLPEDKILETF 720

; Sequence 10, Application US/10379981
; Publication No. US20040001820A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pharmaceuticals, Inc.
; APPLICANT: Hahn, Chang S
; APPLICANT: Liu, Hong S
; TITLE OF INVENTION: HUMAN DEUBIQUITINATING PROTEASE GENE ON CHROMOSOME 7 AND ITS
; FILE REFERENCE: USAV2002-0022 WO PCT
; CURRENT APPLICATION NUMBER: US/10/379,981
; PRIOR FILING DATE: 2003-03-05
; PRIOR APPLICATION NUMBER: GB 0218518.9
; PRIOR FILING DATE: 2002-09-08
; PRIOR APPLICATION NUMBER: US 60/366,601
; PRIOR FILING DATE: 2002-03-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 10
; LENGTH: 1316
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-379-981-10

Query Match          99.0%; Score 4028; DB 12; Length 1251;
Best Local Similarity 99.7%; Pred. No. 5.5e-286;
Matches 767; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTIVDKASSSDPSAYONQPGSEAVSPGDMAGSAGWAVSSLNDVSNHTLSLGPVPGA 60
DB 1 MTIVDKASSSDPSAYONQPGSEAVSPGDMAGSAGWAVSSLNDVSNHTLSLGPVPGA 60
QY 61 VVYSSSVDPKSPKQDQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
DB 61 VVYSSSVDPKSPKQDQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTTPPLANYMLSHEHSKTCCHAEFCMCTQAHITQALSNPGDVVKPMFVI 180
DB 121 FANAALQCLTYTTPPLANYMLSHEHSKTCCHAEFCMCTQAHITQALSNPGDVVKPMFVI 180
QY 181 NEMRRIARHLRFNGQDAHEFLQYTVDMQKACLNGSNKLDHRTQATTLVCOIFGGYLR 240
DB 181 NEMRRIARHLRFNGQDAHEFLQYTVDMQKACLNGSNKLDHRTQATTLVCOIFGGYLR 240
QY 241 RVKCLNCKGVSDTFDPLVDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSCKCKMWPA 300
DB 241 RVKCLNCKGVSDTFDPLVDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSCKCKMWPA 300
QY 301 SKRFTIHRSSNVLTLSLKRFPANFTGGKIAKDVKYPEYLDIRPYMSQNGEPIVVVLAVL 360
DB 301 SKRFTIHRSSNVLTLSLKRFPANFTGGKIAKDVKYPEYLDIRPYMSQNGEPIVVVLAVL 360
QY 361 VHTGFNCHAGHYFCYIKASNGLWTQMNDISVSTSDIRSVLSQQAAYLVFIYRSHDVKNNGE 420
DB 361 VHTGFNCHAGHYFCYIKASNGLWTQMNDISVSTSDIRSVLSQQAAYLVFIYRSHDVKNNGE 420
QY 421 LTHPTHSPGQSSPRPVISQSVVTKNQAAAGFIGPQLPSHMIKNPPLHNGTGPKLDTPESS 480
DB 421 LTHPTHSPGQSSPRPVISQSVVTKNQAAAGFIGPQLPSHMIKNPPLHNGTGPKLDTPESS 480
QY 481 MSSPNGNSVNRASFPVNASAVQNSVNRSSVIEHPKKQKITIISIHKLPRVQCOSQPN 540
DB 481 MSSPNGNSVNRASFPVNASAVQNSVNRSSVIEHPKKQKITIISIHKLPRVQCOSQPN 540
QY 541 LHSNSLENPTKVPSPSTITNSAVOSTSNASTMSVSSKVTKEIPRSESCSQPMNGKSKLN 600
DB 541 LHSNSLENPTKVPSPSTITNSAVOSTSNASTMSVSSKVTKEIPRSESCSQPMNGKSKLN 600
QY 601 SSVLPVYGAESSEDSDESKGLGKENGIGTIVSSHSPQDQDADEEATPHELQEPMTLNGA 660
DB 601 SSVLPVYGAESSEDSDESKGLGKENGIGTIVSSHSPQDQDADEEATPHELQEPMTLNGA 660
QY 661 NSADSDDPKENGLAPDGASCOGQPALHSENPFPAKANGLPGLMPAPLLSLPEDKILETF 720
DB 661 NSADSDDPKENGLAPDGASCOGQPALHSENPFPAKANGLPGLMPAPLLSLPEDKILETF 720
QY 721 RLSNKLKSTDEMSAPGAERGPPEDRDAPQPGSPAESLEPDAAASL 769
DB 721 RLSNKLKSTDEMSAPGAERGPPEDRDAPQPGSPAESLEPDAAAGL 769

RESULT 2
US-10-379-981-9
```

Db 661 NSADSDSPKENGAPDASGCGQFALHSENPFKANGLPGLKMPAPLLSLPDKILETF 720
QY 721 RLSNKLKSTDEMSAPGAEPPEDRDAPQPSGAASLEPPDAASL 769
Db 721 RLSNKLKSTDEMSAPGAEPPEDRDAPQPSGAASLEPPDAAGL 769
RESULT 3
US-10-163-547-1
; Sequence 1, Application US/10163547
; Publication No. US200303037350A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; APPLICANT: Myers, Rachel
; APPLICANT: Rudolf-Owen, Laura A.
; TITLE OF INVENTION: No. US200303037350A1el nucleic acid sequences encoding a
; TITLE OF INVENTION: human ubiquitin protease, lipase, dynamin, short chain
; TITLE OF INVENTION: dehydrogenase, and ADAM-Ts Metalloprotease and uses
; TITLE OF INVENTION: therefore
; FILE REFERENCE: 35800/247836(5800-243)
; CURRENT APPLICATION NUMBER: US/10/163,547
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 09/407,356
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 09/704,918
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: 09/435,311
; PRIOR FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: 09/796,100
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/185,503
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: 09/781,598
; PRIOR FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/182,009
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: 09/782,952
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,408
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 09/496,005
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 762
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-163-547-1
Query Match 98.3%; Score 3999; DB 14; Length 762;
Best Local Similarity 99.9%; Pred. No. 3.5e-284;
Matches 761; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MTIVDKASESSDPSAYONQPGSSEAVSPGMDAGSAGWAVSSINDVSNHTLSLGPVPGA 60
Db 1 MTIVDKASESSDPSAYONQPGSSEAVSPGMDAGSAGWAVSSINDVSNHTLSLGPVPGA 60
QY 61 VVYSSSVDPKSPKSPKQKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
Db 61 VVYSSSVDPKSPKSPKQKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTTPFLANMYLSHEHSKTCCHAEFCMCTMQAHTQALSNPDGVIKPMFVI 180
Db 121 FANAALQCLTYTTPFLANMYLSHEHSKTCCHAEFCMCTMQAHTQALSNPDGVIKPMFVI 180
QY 181 NEMRRIARLRFNGQDAHEFLQYTVDMQKACLNGSNKLDREHTQATTLVCOIFGGYLR 240
Db 181 NEMRRIARLRFNGQDAHEFLQYTVDMQKACLNGSNKLDREHTQATTLVCOIFGGYLR 240
QY 241 RVKCLNCKGVSDTFDPYLDITLIEIKAAQSVNKALEQFVKPEQLDGENSYKSKCKMWPA 300
Db 241 RVKCLNCKGVSDTFDPYLDITLIEIKAAQSVNKALEQFVKPEQLDGENSYKSKCKMWPA 300

Db 241 RVKCLNCKGVSDTFDPYLDITLIEIKAAQSVNKALEQFVKPEQLDGENSYKSKCKMWPA 300
QY 301 SKRFTIHRSSNVLTLSLKRFAFTGKIAKVDKPYEYLDIRPYMSQPNGEPIVYLYAVL 360
Db 301 SKRFTIHRSSNVLTLSLKRFAFTGKIAKVDKPYEYLDIRPYMSQPNGEPIVYLYAVL 360
QY 361 VHTGENCHAGYFCYIKASNGLIWQNDISVSTSDIRSVLSQOAYVLFYIRSHDVKNNGE 420
Db 361 VHTGENCHAGYFCYIKASNGLIWQNDISVSTSDIRSVLSQOAYVLFYIRSHDVKNNGE 420
QY 421 LTHPTSPGQSSPRPVSQRVVTNNKQAAAPGFTGPOLPSHMIKNPHEHNGTGLKDPSSS 480
Db 421 LTHPTSPGQSSPRPVSQRVVTNNKQAAAPGFTGPOLPSHMIKNPHEHNGTGLKDPSSS 480
QY 481 MSSPNGNSSVNRASPVNASVQNVMSVNRSSVPEHPKKQKITTISIHNKLPVRQCSQPN 540
Db 481 MSSPNGNSSVNRASPVNASVQNVMSVNRSSVPEHPKKQKITTISIHNKLPVRQCSQPN 540
QY 541 LHSNSLENPTKPVFSSITINSVOSTSNASTMSVSSKVTKPIRSESCSQPMVNGSKSLN 600
Db 541 LHSNSLENPTKPVFSSITINSVOSTSNASTMSVSSKVTKPIRSESCSQPMVNGSKSLN 600
QY 601 SSVLPVYGAESSESDSDSKGLKENGIGTIVSSHSPGQDAEDEEATPHELOEPMTINGA 660
Db 601 SSVLPVYGAESSESDSDSKGLKENGIGTIVSSHSPGQDAEDEEATPHELOEPMTINGA 660
QY 661 NSADSDSPKENGAPDASGCGQFALHSENPFKANGLPGLKMPAPLLSLPDKILETF 720
Db 661 NSADSDSPKENGAPDASGCGQFALHSENPFKANGLPGLKMPAPLLSLPDKILETF 720
QY 721 RLSNKLKSTDEMSAPGAEPPEDRDAPQPSGAASLEPPDAASL 762
Db 721 RLSNKLKSTDEMSAPGAEPPEDRDAPQPSGAASLEPPDAAGL 762
RESULT 4
US-10-379-981-12
; Sequence 12, Application US/10379981
; Publication No. US20040001820A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pharmaceuticals, Inc.
; APPLICANT: Liu, Hong S
; TITLE OF INVENTION: HUMAN DEUBIQUITINATING PROTEASE GENE ON CHROMOSOME 7 AND ITS
; FILE OF INVENTION: MURINE ORTHOLOG
; FILE REFERENCE: USAV2002-0022 WO PCT
; CURRENT APPLICATION NUMBER: US/10/379,981
; CURRENT FILING DATE: 2003-03-05
; PRIOR APPLICATION NUMBER: GB 0218518.9
; PRIOR FILING DATE: 2002-09-08
; PRIOR APPLICATION NUMBER: US 60/366,601
; PRIOR FILING DATE: 2002-03-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 12
; LENGTH: 1326
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-379-981-12
Query Match 78.6%; Score 3198.5; DB 15; Length 1326;
Best Local Similarity 80.1%; Pred. No. 3.9e-225;
Matches 620; Conservative 50; Mismatches 87; Indels 17; Gaps 7;
QY 1 MTIVDKASESSDPSAYONQPGSSEAVSPGMDAGSAGWAVSSINDVSNHTLSLGPVPGA 60
Db 1 MTIVDK-TFSPDSTCQNPFGSCEAVSPEDMDTGSASGAVSSISDVSSHLPGLPVPGA 59
QY 61 VVYSSSVDPKSPKSPKQKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
Db 61 VVYSSSVDPKSPKSPKQKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTTPFLANMYLSHEHSKTCCHAEFCMCTMQAHTQALSNPDGVIKPMFVI 180
Db 121 FANAALQCLTYTTPFLANMYLSHEHSKTCCHAEFCMCTMQAHTQALSNPDGVIKPMFVI 180

```
Db 120 FANAALQCLTYTPPLANTMLSHEHSKTCBAGFCMCTMQTHITQALSNPQGVIVKPMFEVI 179
QY 181 NEMRIARHLRFQNGQEDAEHFLQYTVDAKQACLNGSNKLDHRTQATTLVQCIFGGYILRS 240
Db 180 NEMRIARHLRFQNGQEDAEHFLQYTVDAKQACLNGSNKLDHRTQATTLVQCIFGGYILRS 239
QY 241 RVKCLNCKGVSDTDFPYLDITLTKAAQSVNKALEBQFVKPKPOLDGENSYKSKCKKQMPA 300
Db 240 RVKCLNCKGVSDTDFPYLDITLTKAAQSVNKALEBQFVKPKPOLDGENSYKSKCKKQMPA 299
QY 301 SKRFTIHRSSNNVLTSLKRFANFTGGKTAKOVKPEYLDIIRYMSOPNGEPTVVYLYAVL 360
Db 300 SKRFTIHRSSNNVLTSLKRFANFTGGKTAKOVKPEYLDIIRYMSOPNGEPTVVYLYAVL 359
QY 361 VHTGPNCHAGHYFCYIKASNGLIWQMNDSIVSTSDIRSVLSQOAYVLYIRSHDVKNNGE 420
Db 360 VHTGPNCHAGHYFCYIKASNGLIWQMNDSIVSTSDIRAVLNQOAYVLYIRSHDVKNNGE 419
QY 421 LTHPTSPGQSSPRPVIQORVVNKAAPGFTGPQLPSHMIKPNPHLNGTGPLKDTFSS 480
Db 420 SAHPAHSPGQSSPRPVGVSQVNNKQVAFGFTGPQLPSHVMKNTPHLNGTTPVKDTPSS 479
QY 481 MSSPNCNSVNRASPVNASASVQNSVNRSSVIBPHPKKOKITITSHNKLPRVQCQSQPN 540
Db 480 VSSPNCNTSVNRASPVNASASVQNSVNRSSVIBPHPKKOKITITSHNKLPRVQCQSQPN 539
QY 541 --LHNSLENPTKPVPSSTITN-SAVQSTSNASTMSVSSKVTKPIPRSSCSQPVNMGKS 597
Db 540 NSLHGFCLEAPSKAPSSITNPSAIQSTSNVPTTSTSP-----SEACPKPMVNGKA 591
QY 598 KLNSSVLPVYGAESSEDSDEESKGLKENGIGTIVSSHS--PGQDAED-EEATPHELQEP 654
Db 592 KYGASVLPVYGAESSEDSDEESKGLAKENGVDMMAGTHSDRPEAAADGAEASSELQEP 651
QY 655 MTLGANSADSDSPKENGGLAPDAGSCQGOALHSENPFAKANGLPKMLPAPLLSLP 714
Db 652 VILGANSADSDS--QENSLAFDSACQOPELHNTENLFSKLNGLPKVTPAPLQSPED 709
QY 715 KILETFLSNKLGSTDEMSAPGAERPPEDRDAEPQGSAPAESLEEDPAAS 768
Db 710 RILETFLSNKLGSTDEMSAPGAERPPEDRDAEPQGSAPAESLEEDPAAS 763

RESULT 5
US-09-972-525-1
; Sequence 1, Application US/09972525
; Publication No. US20020086407A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 23484, A NOVEL HUMAN UBIQUITIN
; FILE REFERENCE: 5800-51
; CURRENT APPLICATION NUMBER: US/09/972,525
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/408,865
; PRIOR FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1123
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-525-1

Query Match 35.0%; Score 1423.5; DB 12; Length 1123;
Best Local Similarity 37.3%; Pred. No. 4.le-95;
Matches 347; Conservative 112; Mismatches 238; Indels 233; Gaps 25;

QY 1 MTIVDKASESSDPSAYQNGQSSPAVSPGDMWDAGSAGVASSINDV----- 47
Db 1 MPIVDKLKEAL-----KPGRKADSDGELGKLILASSAKKVLQKTEFPASKSFYQL 53
QY 48 ----SNHTLSLGPVPGAVVY-SSSSVPDKSPKQDQALGDGIAPQKVLFTSEKICLK 102
```

```
Db 54 EALKSKYVLLNPKTEGASRHSKSGDPPARRQSGSEHTYESCGDVPAPQKVLFPPTLSLR 113
QY 103 WOQTHRVGAGLONGNTCFANAALQCLTYTPPLANTMLSHEHSKTCBAGFCMCTMQAH 162
Db 114 WERVFRVAGLHNLGNTCFNLNATIQCLTYTPPLANTMLSHEHSKTCBAGFCMCTMQAH 173
QY 163 ITOALSNGPDVIMKMFVINEMRRIARHLRFQNGQEDAEHFLQYTVDAKQACLNGSNKLD 222
Db 174 IVQAFANSNAIKPVFIRDILKKIARHFRFQNGQEDAEHFLQYTVDAKQACLNGCAKLD 233
QY 223 HTQATTLVQCIFGGYILRSRVKCLNCKGVSDTDFPYLDITLTKAAQSVNKALEBQFVK 282
Db 234 QTQATTLVQIFGGYILRSRVKCSVCKSVSDTDFPYLDVALEIRQAANIVRALELVKAD 293
QY 283 LDGENSYKSKCKKQMPVPAKSKFTIHRSSNNVLTSLKRFANFTGGKTAKOVKPEYLD 342
Db 294 LSGENAYNCAKCKKVPASKKFTIHRSSNNVLTSLKRFANFTGGKTAKOVKPEYLD 353
QY 343 YMSQNGEPIVVVLYAVLVHTGFNCHAGHYFCYIKASNGLIWQMNDSIVSTSDIRSVLS 402
Db 354 YMSQNGEPIVVVLYAVLVHTGFNCHAGHYFCYIKASNGLIWQMNDSIVSTSDIRSVLS 413
QY 403 QAYVLYIRSHDVKNNGE--LTHPTSPGQSSPRP-----VLSQRVVTKQ-- 446
Db 414 QAYVLYIRSHDVKNNGE--LTHPTSPGQSSPRP-----VLSQRVVTKQ-- 473
QY 447 -----AAGFTGPQLPS-----HMIKPNPHLNGTGPLK 475
Db 474 GTMKXPHTTEIGVP--SRNGSTLGLKQNGCIPPKLPSPGSPKLSQVTHM---PTILD 530
QY 476 TPSSSMSSP-----NGNSS-----VNRASPV----- 496
Db 531 DPGKVKVPAPPOHFPSPRTAQLPGTSNNSNRSGSQROGQSDRDVVLSTSPKLLATAT 590
QY 497 -----NASASVQNSVNRSSVIBPHK-----KOKITISI--H 527
Db 591 ANGHGLKGNDESAGLDRRGSSSSSPHSASDSKAPQTPRSGAAHLCDSQETNCSTAGH 650
QY 528 NKLPVRQCSQ-----PNLHNSLE-----NPT 550
Db 651 SKTPPSGADSKTVKLKSPVLSNTTTPASTMTGPPPAKKALSAKASTIWRATGNDRPP 710
QY 551 KPVPSSITNSAVQSTSN--ASTMSV-SSKVTKPIPRSESCSQPYMNGKSKLNSVLVP 606
Db 711 PRSPSSDLTHP--MKTSHPVVASTWVHRARAVSPAPQSSSRLOPPFSPHPTLLSSTPK 768
QY 607 YGAESSEDSDEESKGLKENGIGTIVSSHS PGQDAEDAEATPHELQEPMTLNGANSAD 666
Db 769 PGTSEPRSCSSISTALPQVN-----EDLVSLPHOLPEA----- 801
QY 667 SDPKENGLAPDAGSCQGO-ALHSENPFAKANGLPKMLPAPLLSPEDKILETFLSNK 725
Db 802 SEPPQSPSEKRRKFTVGEFQRLGSE-----TFLPOHIREA--TAAPHGK-----RKRK 847
QY 726 LKGSTDEMSAPGAERPPEDRDAEPQGS 755
Db 848 KKKRPEDTAASALQEG-----QTQRPQGS 872

RESULT 6
US-10-371-905A-20
; Sequence 20, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
```

; SEQ ID NO 20
 ; LENGTH: 565
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: MISC FEATURE
 ; OTHER INFORMATION: human DUB4.8
 US-10-371-905A-20

Query Match 28.3%; Score 1150.5; DB 12; Length 565;
 Best Local Similarity 46.7%; Pred. No. 1.6e-75;

Matches 244; Conservative 78; Mismatches 162; Indels 39; Gaps 8;

```

QY 28 PCDMDAGSASGAVSSLDVSNHTLSLGPVPCAVVYSSSSVDPKSPQKQDQALGDGIA 87
D 1 MEDDLSYLGGGEQFNHFSKLT--SSRPDAFAFIEIQTSLPEKSPFLSCARVDLCDDLA 91
QY 88 PPKVLPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYMLSHSHT 147
D 92 FVARQAPREKLPPLSSRRPAAVAGLQNGMNTCYVNASLQCLTYTPPLANYMLSHSHT 151
QY 148 CHABGFCMCTMQAHITCALSNPGDVIPKMFVINEMRRIARHLRGNQEDAEHEFLQY 207
D 152 CHRHGCMCTMQAHITRALNPHVITP-----SQALAAAGFHRGKQEDAEHEFLMFTVD 205
QY 208 AMQKACLSNKLDRHTQATTLVCOIFGGYLSRVKLCNCKGVSDTFDPYLDITLLEIKAA 267
D 206 AMKACLPKGKQVDHHSKDTTLHQIFGGYWSQIKLCHCHGISDTPDPYLDIALDIOA 265
QY 268 OSVNAKALROFVKPQDGENSYKCKCKMVPASKRFTIHRSSNVLTSLKRFANFTGGK 327
D 266 OSVQALQELVKPELNGENAYHCGVCLQAPASKTTLHTSAKVLILVLRKFSVDVTGN 325
QY 328 IAKDVKYPEYLDIRPYMSQNGEPIVYVAVLVHTGFNCHAGHYFCYIKASNLGLWQMN 387
D 326 IAKNVQPECLDMQPYMSQQNGTGLVYVAVLVHTGFNCHAGHYFCYIKASNLGLWQMN 385
QY 388 DSIVSTSDIRSVLSQQAYVLYFIR-----SHDVKNKGELTHPTSPGSSPRPVISOR 440
D 386 DAEVTAASITSVLSQQAYVLYFIOKSEWERSHESVSRGE-----PRALGAED 433
QY 441 VTNKQAPGFI-----GPQLPSHMIKNPPLHNGTGPLKPTPSSMSPNNGN--SVN 491
D 434 --TDRRATQELKRDHPCLQAPDELHVERATQESTLDHWKFLQEQNKTKPEFNVKVE 491
QY 492 RASPVNASASVQNSVNRSSVPIHPKQKOKITISIHKNLIPVRQ 534
D 492 GTLP--PDVLVHQSRYKCGKGNHHPQEQSSLLNLSSTPTHQ 532
  
```

RESULT 7

US-10-107-695B-2
 ; Sequence 2, Application US/10107695B
 ; Publication No. US20030022201A1
 ; GENERAL INFORMATION:

; APPLICANT: Kapeller-Libermann, Rosana
 ; TITLE OF INVENTION: 68999, Human Ubiquitin
 ; FILE REFERENCE: Carboxyl-Terminal Hydrolase Family Member and Uses Therefor
 ; CURRENT APPLICATION NUMBER: US/10/107,695B
 ; PRIOR FILING DATE: 2002-09-10
 ; PRIOR APPLICATION NUMBER: 60/279,184
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2
 ; LENGTH: 530
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens

US-10-107-695B-2

Query Match 28.2%; Score 1148; DB 14; Length 530;
 Best Local Similarity 43.1%; Pred. No. 2.2e-75;

Matches 252; Conservative 89; Mismatches 164; Indels 80; Gaps 12;

```

QY 31 MDAGSASGAVSSLDVSNHTLSLGPVPCAVVYSSSSVDPKSPQKQDQALGDGIA 90
D 1 MEDDLSYLGGGEQFNHFSKLT--SSRPDAFAFIEIQTSLPEKSPFLSCARVDLCDDLA 99
QY 91 KVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYMLSHSHT 150
D 60 RQAPREKLPPLSSRRPAAVAGLQNGMNTCYVNASLQCLTYTPPLANYMLSHSHT 119
QY 151 EGFCMCTMQAHITCALSNPGDVIPKMFVINEMRRIARHLRGNQEDAEHEFLQY 210
D 120 HKGCMCTMQAHITRALNPHVITP-----SQALAAAGFHRGKQEDAEHEFLMFTVD 173
QY 211 KACLSNKLDRHTQATTLVCOIFGGYLSRVKLCNCKGVSDTFDPYLDITLLEIKAA 270
D 174 KACLPKGKQVDHHSKDTTLHQIFGGYWSQIKLCHCHGISDTPDPYLDIALDIOA 233
QY 271 NKALROFVKPQDGENSYKCKCKMVPASKRFTIHRSSNVLTSLKRFANFTGGK 330
D 234 QQALQELVKPELNGENAYHCGVCLQAPASKTTLHTSAKVLILVLRKFSVDVTGN 293
QY 331 DVKYPEYLDIRPYMSQNGEPIVYVAVLVHTGFNCHAGHYFCYIKASNLGLWQMN 390
D 294 NVQPECLDMQPYMSQQNGTGLVYVAVLVHTGFNCHAGHYFCYIKASNLGLWQMN 353
QY 391 VSTSDIRSVLSQQAYVLYFIR-----SHDVKNKGELTHPTSPGSSPRPVISOR 443
D 354 VTASITSVLSQQAYVLYFIOKSEWERSHESVSRGE-----PRALGAED--T 399
QY 444 NKQAPGFI-----GPQLPSHMIKNPPLHNGTGPLKPTPSSMSPNNGN--SVN 495
D 400 DRRATQELKRDHPCLQAPDELHVE-----RATQESTLDHWKFLQEQNKTKP 448
QY 496 VNASASVQNSVNR-----SSVPIHPKQKOKITISIHKNLIPVRQ 534
D 449 -----EFNVKVEGTLPPDVLVHQSRYKCGKGNHHP-----EQSSLLNLSST 494
QY 550 TKVPVSTITNSAVQSTNSATM--SVSSKVTKPIRSESCSPVM 593
D 495 TD-----QESMNTGTLASLQGRTRRAKGNKCKRSIL 527
  
```

RESULT 8

US-10-371-905A-16
 ; Sequence 16, Application US/10371905A
 ; Publication No. US20030224969A1
 ; GENERAL INFORMATION:

; APPLICANT: Hahn, Chang
 ; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
 ; FILE REFERENCE: USAV2001/0185
 ; CURRENT APPLICATION NUMBER: US/10/371,905A
 ; CURRENT FILING DATE: 2003-02-20
 ; NUMBER OF SEQ ID NOS: 54
 ; SOFTWARE: Patent in version 3.1

; SEQ ID NO 16
 ; LENGTH: 530
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens

; NAME/KEY: MISC FEATURE
 ; OTHER INFORMATION: human DUB4.6

US-10-371-905A-16

Query Match 28.1%; Score 1142.5; DB 12; Length 530;
 Best Local Similarity 46.7%; Pred. No. 5.5e-75;

Matches 243; Conservative 77; Mismatches 161; Indels 39; Gaps 8;

```

QY 31 MDAGSASGAVSSLDVSNHTLSLGPVPCAVVYSSSSVDPKSPQKQDQALGDGIA 90
D 1 MEDDLSYLGGGEQFNHFSKLT--SSRPDAFAFIEIQTSLPEKSPFLSCARVDLCDDLA 99
  
```

Query Match	28.0%;	Score 1140.5;	DB 12;	Length 530;	
Best Local Similarity	46.5%;	Pred. No. 7.7e-75;			
Matches 242;	Conservative	78;	Mismatches 161;	Indels 39;	Gaps 8;
QY	31	MDAGSASGAVSSLNDVSNHTLSLSPGVFEGAVYVYSSSSVPDKSPQKDQALGDIAPPQ	90		
DB	1	MEDDSLVLGGEWQFNHFSKLT-SRRPDAAFAEIIQRTSLPEKSPSLCETRVLDCLDLPFA	59		
QY	91	KVLPESEKILCKWQOTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYMLSHESKTCCHA	150		
DB	60	QLAPREKLPILSSRRPAAGVAGLQNMGNMTCYNASLQCLTYTPPLANYMLSREHSQTCRR	119		
QY	151	BGFQMCWTQWAIHQALSNPGDVIKPMFVINEMRRIRARHLPFGQEDAEHLQYTVDAQQ	210		
DB	120	HKGCMCLCTQWAIHQALSNPGDVIKPMFVINEMRRIRARHLPFGQEDAEHLQYTVDAQQ	173		
QY	211	KACLLNGSNKLRDHRTQATTLVCOIFGGYLRSRVKLNCNLKGVSDTFDPYDLDTLTETKAAQSV	270		

Query Match	28.3%;	Score 1139.5;	DB 12;	Length 574;
Best Local Similarity	46.5%;	Pred. No. 1e-74;		
Matches	243;	Conservative 77;	Mismatches 164;	Indels 39; Gaps 8;
QY	28	PGDMDAGSAGWVSSLNDVSNHHTLSLGPVPGAVVYSSSVDPKSKSPQKDQALGDGIA	87	
DB	33	PVDMEEDSLYLGGEWQFNHFSKLT-SRPDAAFAEIORTSLPEKSPISCETRVLDLDDLA	91	
QY	88	PPQKVLPPSEKICLKWOOTHVRVAGLQMLGNMTCFANAALQCLTTPPLANTMLSHEHSKT	147	
DB	92	PVARQLAPPEKLPISNRRPPAAVAGLQNMGNMTCVNASLQCLTTPPLANTMLSREHSQT	151	
QY	148	CHABGFCMCTMQAHIITQALSNPGDVTIKPMFVIMEMRIARHLRFQNGQDAHEFLQYITVD	207	
DB	152	CHRHKGCMCTMQAHIITRALNHPGHVQP-----SQALAAFGPHRGQEDAEHEFLMFTVD	205	
QY	208	AMQKACLINGSNKLDRHTQATTLVCQIFGGYLRSRVKCLNCKGVSDTFPPYLDITLLEIKAA	267	
DB	206	AMKCACLPGHKQVDHHSKDTLLHQIFGGYWRSQIKLHCHGISDTFPPYLDIALDIOAA	265	
QY	268	QSVNKAALQFVKPECLDGENSYKCSCKKQWVPASKRFTTHRSSNVLTLSLKRANFTGGK	327	
DB	266	QSVQQAHLQVKPEELNGENAYHCGVCLQAPASKTLLTHTSAKVLILVLRFSFVDTGNK	325	
QY	328	TAKDVKYPEYDLDIRYNSQNGPEPIVYLYAVLHVHTGENCHAGHVFCYIKASNGWLGYOWN	387	
DB	326	IDKNVQYPECLDMKLYNQTSNGPLVYLYAVLVHAGWSCHNGHVSFYVIAQEGQWYKMD	385	

Query Match	28.3%;	Score 1139.5;	DB 12;	Length 574;
Best Local Similarity	46.5%;	Pred. No. 1e-74;		
Matches	243;	Conservative 77;	Mismatches 164;	Indels 39; Gaps 8;
QY	28	PGDMDAGSAGWVSSLNDVSNHHTLSLGPVPGAVVYSSSVDPKSKSPQKDQALGDGIA	87	
DB	33	PVDMEEDSLYLGGEWQFNHFSKLT-SRPDAAFAEIORTSLPEKSPISCETRVLDLDDLA	91	
QY	88	PPQKVLPPSEKICLKWOOTHVRVAGLQMLGNMTCFANAALQCLTTPPLANTMLSHEHSKT	147	
DB	92	PVARQLAPPEKLPISNRRPPAAVAGLQNMGNMTCVNASLQCLTTPPLANTMLSREHSQT	151	
QY	148	CHABGFCMCTMQAHIITQALSNPGDVTIKPMFVIMEMRIARHLRFQNGQDAHEFLQYITVD	207	
DB	152	CHRHKGCMCTMQAHIITRALNHPGHVQP-----SQALAAFGPHRGQEDAEHEFLMFTVD	205	
QY	208	AMQKACLINGSNKLDRHTQATTLVCQIFGGYLRSRVKCLNCKGVSDTFPPYLDITLLEIKAA	267	
DB	206	AMKCACLPGHKQVDHHSKDTLLHQIFGGYWRSQIKLHCHGISDTFPPYLDIALDIOAA	265	
QY	268	QSVNKAALQFVKPECLDGENSYKCSCKKQWVPASKRFTTHRSSNVLTLSLKRANFTGGK	327	
DB	266	QSVQQAHLQVKPEELNGENAYHCGVCLQAPASKTLLTHTSAKVLILVLRFSFVDTGNK	325	
QY	328	TAKDVKYPEYDLDIRYNSQNGPEPIVYLYAVLHVHTGENCHAGHVFCYIKASNGWLGYOWN	387	
DB	326	IDKNVQYPECLDMKLYNQTSNGPLVYLYAVLVHAGWSCHNGHVSFYVIAQEGQWYKMD	385	


```
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1021
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.2a
US-10-371-905A-8

Query Match      27.7%; Score 1128.5; DB 12; Length 1021;
Best Local Similarity 46.1%; Pred. No. 1.5e-73;
Matches 241; Conservative 78; Mismatches 165; Indels 39; Gaps 8;

QY      28  PGDMAGSAGVSSLDVNSHTLSLGPVPGAVVYSSSSVDPKSPKQDQALGDGIA 87
DB      489  PVDMEDDSLYLGGEWQFNHFSKLT--SSRDAFAEIQTSLSEKPLSCETRVDLCDLLA 547

QY      88  PPQKVLFFSEKICLKQWQTHRVGAGLNGNTCFANAALQCLTYTTPPLANYMLSHEHST 147
DB      548  PVARQLAPREKPLPSRRRAAAGAGLQNGNTCYVNASLQCLTYTTPPLANYMLSREHSQT 607

QY      148  CHAEGFCMMCTQAHITQALSNGPDIKPMFVINEMRRTARHLRFGNQDAHEFLQYTD 207
DB      608  CHRHGCMCTMQAHITRALHNPCHVIOQ-----SQALAAAGFHRGQDAHEFLMFTVD 661

QY      208  AMOKACLGNSKLDRTQATTLVQIFGVLRSVKLCNCKGVSDTFDPIYDLITLLEIKAA 267
DB      662  AMKACLPQGHQVLPPEELNGENAYHSGVCLQRAPASKTLTLTSAKVLILVLRKPSFSDVTGNK 781

QY      268  QSNKALEQVFKPEQLDGENSKYCKSKCKKMPASKRFTTHRSSNVLTLSLKRFANFTGGK 327
DB      722  OSVQALEQLVXPEELNGENAYHSGVCLQRAPASKTLTLTSAKVLILVLRKPSFSDVTGNK 781

QY      328  IAKDVKPYEYLDIRPMSQNGEPIYVLYVHTGFNCHAGHYCYIKASNLWYQWN 387
DB      782  IAKNVQYPECLDMQPYMSQNGTGLVYVLYVHTGFNCHAGHYCYIKASNLWYQWN 841

QY      388  DSTVSTDIRSVLSQAYVLYFYIR-----SHDVKNNGGELTHPTSPGSSPRPVISQR 440
DB      842  DAEVTAASITSAUSQAYVLYFYIQSEWERHSESVSRGRE-----PRALGTED 889

QY      441  VVTNKQAAPGFI-----GPQLPSHMIKNPHLNGTGLKDTPTSSSMSPNGN--SSVN 491
DB      890  --TDRATQGLKRDHPCLQAPLDELHVERATQESTLDHWKFLQEQNKTKPFNVKVE 947

QY      492  RASPVNASASVQWNSVNRSSVIEHPKQKQITSIHNKLPVRQ 534
DB      948  GTLP--PDVLVIHQSKYKCGMKNHHPQSQSSLLNLSSTPTHQ 988

RESULT 14
US-10-371-905A-24
; Sequence 24, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 24
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human LUB4.11
US-10-371-905A-24

Query Match      27.7%; Score 1126.5; DB 12; Length 633;
Best Local Similarity 44.6%; Pred. No. 1.1e-73;
Matches 246; Conservative 84; Mismatches 169; Indels 53; Gaps 11;

QY      1  MTIV--DKASESSDPEAYQNQPGSEAVSPGMDAGSASWGAVSSLDVNSHTLSLGPVP 58
DB      84  LTLRSEASALSSCFEVAERMAAAKI---DMEDDSLYLGGEWQFNHFSKLT--SSRPA 138

QY      59  GAVVYSSSSVDPKSKIPKQDQALGDGIAPPOKVLFPSEKICLKQWQTHRVGAGLQNLN 118
DB      139  APAEIQTSLSEKIS-----PLSYDLAPVARQLAPREKPLPSRRRAAAGLQNMGN 190

QY      119  TCFANAALQCLTYTTPPLANYMLSHEHSTCHAEFGCMCTMQAHITQALSNGPDIKPMF 178
DB      191  TCYVNASLQCLTYTTPPLANYMLSREHSQTCRHKGCMCTMQAHITRALHNPCHVIOQ-- 248

QY      179  VINEMRRTARHLRFGNQDAHEFLQYTVDMQKACLGNSKLDRTQATTLVQIFGGYL 238
DB      249  ----SQALAAAGFHRGQDAHEFLMFTVDAMKACLPKHKQYDHHKDTTLTHQIFGGY 304

QY      239  RSRVKLCNCKGVSDTFDPIYDLITLLEIKAAQSVNKALEQVFKPEQLDGENSKYCKSKCKMV 298
DB      305  RSOIKCLCHGISTDFDPIYDLITLLEIKAAQSVNKALEQVFKPEQLDGENSKYCKSKCKMV 364

QY      299  PASKRFTTHRSSNVLTLSLKRFANFTGGKIAKDVKPYEYLDIRPMSQNGEPIYVLYA 358
DB      365  PASKTLTLTSAKVLILVLRKPSFSDVTGNKIAKDVKPYEYLDIRPMSQNGEPIYVLYA 424

QY      359  VLVHTGFNCHAGHYCYIKASNLWYQWNSIVSTDIRSVLSQAYVLYFYIR----- 411
DB      425  VLVHAEWSCHNGHYCYIKASNLWYQWNSIVSTDIRSVLSQAYVLYFYIQSEWERH 484

QY      412  SHDVKNNGGELTHPTSPGSSPRPVISQSVTNKQAAPGFI-----GPQLPSHMIKN 463
DB      485  SESVSRGRE-----PRALGAED--TDRATQGLKRDHPCLQAPLDELHVER 530

QY      464  PHLNLTGPKDTPTSSSMSPNGN--SSVNRAASVNASASVQWNSVNRSSVIEHPKQKQI 522
DB      531  ATQESTLDHWKFLQEQNKTKPFNVKVKGTLP--PDVLVIHQSKYKCGMKNHHPQSQSS 588

QY      523  TSIHNKLPVRQ 534
DB      589  LNLSSSTPTHQ 600

RESULT 15
US-10-408-765A-2752
; Sequence 2752, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenna, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
```



```
; SEQ ID NO 2752
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2752

Query Match      27.6%; Score 1125; DB 16; Length 472;
Best Local Similarity 50.9%; Pred. No. 8.9e-74;
Matches 239; Conservative 63; Mismatches 126; Indels 42; Gaps 9;

Qy 49 NHTLSL-GPVGAVV--YSSSSVPDKSPKQKQALGDGIAPPQKVLFPSEKICLKWQQ 105
Db 15 NHFSKLTSPRDAAFAEIQRTSLPEKSPSCSETRVDLCYLAQVAPQALPREKLPSSRR 74
Qy 106 THRYGAGLQNLGNTCFANAALQCLTYTPPLANYMLSHHSKTCBAEGFCMMCTWOAHITQ 165
Db 75 PAAVAGLQNMGNTCYVNASLQCLTYTPPLANYMLSHHSQTCRHRKGCMLCTNQAHITR 134
Qy 166 ALSNPGDVIKPMFVINEMRRIARHLRFQGNQEDAEHEFLQYTVDMOKACLNGSNKLDHRHQ 225
Db 135 ALHNPCHVIQF-----SQALAGFHRGKQEDAEHEFLMTVDAMKACLPQHKQVDHHSK 188
Qy 226 ATTIVCOIFGGLRSRVKCLNCKGVSTDFDYLITLTKAAQSVNKALEQFVKPEQLDG 285
Db 189 DTLIHQIFGGYWRSQIKCLHCHGISDTPDYLIDIALDIQAAQSVQQALEQLVKPEELNG 248
Qy 286 ENSYKCSKCKKMWPKSRFTIHRSSNVLTLSLKRFANFTGKIAKDKYKPYLDIRPYMS 345
Db 249 ENAYHCGVCLQRPASKTLTLHTSAKVLILVKKRFSQVTKGKIKNVQYPECLDMQPYMS 308
Qy 346 QPNGEPIVYVLYAVLVHTGFNCHAGHYFCYIKASNGLWYQWNDISVSTDIRSVLSQAY 405
Db 309 QQNTGPLVYVLYAVLVHAGSCHNGHYFSYVKAQEGQWKIDDAEVTASSITSVLTQAY 368
Qy 406 VLPYIRSHDVANGEL-THPTSPGQSSPRPVI-----SQRVVTKQ-----AA 448
Db 369 VLFYIQSEWERHKSQVETVNP---SVRLSLFLFTQNSQFFLTQKQISIMEKKFSISG 425
Qy 449 PGFIGPQLPSHMIKNPPLNGTGPLK-----DTPSSSMSPNGNSVNR 492
Db 426 PAQARP-----SNPCHSLCGLGQRRGCGSPGSGTRGPEGLSLRR 468
```

Search completed: August 10, 2004, 20:05:45
Job time : 53 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 20:13:12 ; Search time 21 Seconds
(without alignments)
3549.925 Million cell updates/sec

Title: US-10-049-745-4
Perfect score: 775
Sequence: 1 MTIVDKASESSDPAYONQP.....AAESLEPPDAAASLFFPSESG 775

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 10

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	1.3	887	2 C86453	CDS protein F9L11.

ALIGNMENTS

RESULT 1
C86453
CDS protein F9L11.5 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 25-Aug-2003
C;Accession: C86453
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: C86453
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-887 <STO>
A;Cross-references: GB:AE005172; NID:96910582; PIDN:AAF31287.1; GSPDB:GN00141
C;Genetics:

A;Gene: F9L11.5
A;Map position: 1
C;Superfamily: ubiquitin C-terminal hydrolase Ubpl2

Query Match 1.3%; Score 10; DB 2; Length 887;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 GLQLNGNTCF 121
|||
Db 297 GLQLNGNTCF 306

Search completed: August 10, 2004, 20:17:08
Job time : 21 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 20:16:43 ; Search time 50 Seconds

(without alignments)
4862.086 Million cell updates/sec

Title: US-10-049-745-4

Perfect score: 775

Sequence: 1 MTVIDKASSESDPSAYQNQP.....AAESLEEDPAASLFPFSEG 775

Scoring table: OIRGO

Gapop 60.0 , Gapext 60.0

Searched: 1291235 seqs, 313682936 residues

Word size : 10

Total number of hits satisfying chosen parameters: 44

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	666	85.9	1251	12	US-10-399-645-7
2	666	85.9	1316	15	US-10-379-981-9
3	661	85.3	762	14	US-10-163-547-1
4	112	14.5	112	9	US-09-864-761-47587
5	80	10.3	1326	15	US-10-379-981-12
6	66	8.5	66	9	US-09-864-761-46742
7	16	2.1	151	12	US-10-371-905A-48
8	16	2.1	248	12	US-10-371-905A-46
9	16	2.1	271	12	US-10-371-905A-44
10	16	2.1	314	15	US-10-274-639-11
11	16	2.1	314	16	US-10-333-574-11
12	16	2.1	357	12	US-10-371-905A-40
13	16	2.1	398	12	US-10-371-905A-42
14	16	2.1	472	16	US-10-408-765A-2752
15	16	2.1	530	10	US-09-372-348-11

16	2.1	530	12	US-10-371-905A-12	Sequence 12, Appl
17	2.1	530	12	US-10-371-905A-16	Sequence 16, Appl
18	2.1	530	12	US-10-371-905A-18	Sequence 18, Appl
19	2.1	530	12	US-10-371-905A-30	Sequence 30, Appl
20	2.1	530	12	US-10-371-905A-34	Sequence 34, Appl
21	2.1	530	14	US-10-107-695B-2	Sequence 2, Appl
22	2.1	565	12	US-10-371-905A-20	Sequence 20, Appl
23	2.1	574	12	US-10-371-905A-14	Sequence 14, Appl
24	2.1	633	12	US-10-371-905A-24	Sequence 24, Appl
25	2.1	1021	12	US-10-371-905A-8	Sequence 8, Appl
26	1.9	497	12	US-10-371-905A-28	Sequence 28, Appl
27	1.8	245	10	US-09-372-348-13	Sequence 13, Appl
28	1.8	748	10	US-09-372-348-15	Sequence 15, Appl
29	1.8	1123	12	US-09-972-525-1	Sequence 1, Appl
30	1.7	360	10	US-09-372-348-9	Sequence 9, Appl
31	1.7	530	12	US-10-371-905A-6	Sequence 6, Appl
32	1.7	530	12	US-10-371-905A-26	Sequence 26, Appl
33	1.7	530	12	US-10-371-905A-32	Sequence 32, Appl
34	1.7	530	12	US-10-371-905A-36	Sequence 36, Appl
35	1.7	1016	12	US-10-371-905A-4	Sequence 4, Appl
36	1.5	150	12	US-10-424-599-243156	Sequence 243156,
37	1.4	530	12	US-10-371-905A-10	Sequence 10, Appl
38	1.4	623	12	US-10-424-599-276474	Sequence 276474,
39	1.4	715	12	US-10-371-905A-22	Sequence 22, Appl
40	1.4	844	16	US-10-437-963-132523	Sequence 132523,
41	1.3	240	12	US-10-424-599-246594	Sequence 246594,
42	1.3	417	12	US-10-371-905A-38	Sequence 38, Appl
43	1.3	549	12	US-10-424-599-223768	Sequence 223768,
44	1.3	1003	12	US-10-425-114-47268	Sequence 47268, A

ALIGNMENTS

RESULT 1

US-10-399-645-7
; Sequence 7, Application US/10399645
; Publication No. US20040029249A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; LEE, Ernestine A.
; APPLICANT: HAFALIA, April J.A.; YUE, Henry
; APPLICANT: LAL, Preeti G.; YAO, Monique G.
; APPLICANT: LU, Yan; CHAWLA, Narinder K.
; APPLICANT: WARREN, Bridget A.; LU, Dying Aina M.
; APPLICANT: BAUGHN, Mariah R.; DELEGERNE, Angelo M.
; APPLICANT: BURFORD, Neil; BOROWSKY, Mark L.
; APPLICANT: LEE, Sally; XU, Yuming
; APPLICANT: GRIFFIN, Jennifer A.; KALLICK, Deborah A.
; APPLICANT: GANDHI, Ameena R.; ARVIZU, Chandra S.
; APPLICANT: ISON, Craig H.; TANG, Y. Tom
; APPLICANT: AZIMZAL, Falda; ELLIOTT, Vicki S.
; APPLICANT: SWARNAKAR, Anita; RAMKUMAR, Jayalaxmi
; APPLICANT: NGUYEN, Daniel B.; TRIBOULEY, Catherine M.
; APPLICANT: LO, Terence P.; AU-YOUNG, Janice K.
; APPLICANT: THANGAVELU, Kavitha; KEARNEY, Liam
; TITLE OF INVENTION: PROTEASES
; FILE REFERENCE: PI-0263 USN
; CURRENT APPLICATION NUMBER: US/10/399,645
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: PCT/US01/51034
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: US 60/241,573
; PRIOR FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US 60/243,643
; PRIOR FILING DATE: 2000-10-25
; PRIOR APPLICATION NUMBER: US 60/245,256
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/248,395
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: US 60/249,826
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/252,303
; PRIOR FILING DATE: 2000-11-20

; PRIORITY APPLICATION NUMBER: US 60/250,981
; PRIORITY FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 1251
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040023249A1 7997881CD1
US-10-399-645-7

Query Match 85.9%; Score 666; DB 12; Length 1251;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 766; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTIVDKASSSDPSAYONQPGSSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVPGA 60
DB 1 MTIVDKASSSDPSAYONQPGSSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVPGA 60
QY 61 VVYSSSVDPKSPKQKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
DB 61 VVYSSSVDPKSPKQKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTTPPLANYMLSHSHSKTCHAEFGCMCTMQAHTTQALSNPGDVVKPMFVI 180
DB 121 FANAALQCLTYTTPPLANYMLSHSHSKTCHAEFGCMCTMQAHTTQALSNPGDVVKPMFVI 180
QY 181 NEMRRITARHLRFNGQDAHEFLQYTVDMQKACLNGSNKLDRTHTQATTLVCOIFGGYLR 240
DB 181 NEMRRITARHLRFNGQDAHEFLQYTVDMQKACLNGSNKLDRTHTQATTLVCOIFGGYLR 240
QY 241 RVKCLNCKGVSTDFDYLITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSCCKKMWPA 300
DB 241 RVKCLNCKGVSTDFDYLITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSCCKKMWPA 300
QY 301 SKRFTTHRSSNVLTSLKRFANFTGKIAKDVKYPEYLDIRPYMSOPNGEPIVYVLYAVL 360
DB 301 SKRFTTHRSSNVLTSLKRFANFTGKIAKDVKYPEYLDIRPYMSOPNGEPIVYVLYAVL 360
QY 361 VHTGFNCHAGHYFCYIKASNGLWYQNDSTVSTDIRSVLSQAYVLYFIRSHDVKNNGE 420
DB 361 VHTGFNCHAGHYFCYIKASNGLWYQNDSTVSTDIRSVLSQAYVLYFIRSHDVKNNGE 420
QY 421 LTHPTSPGQSSPRPVISQVVTNKQAAFGFTGQPLPSHMIKNPPLHNGTGPLKDTFSSS 480
DB 421 LTHPTSPGQSSPRPVISQVVTNKQAAFGFTGQPLPSHMIKNPPLHNGTGPLKDTFSSS 480
QY 481 MSSPNGSNVNRASPVNASVONWNSVSSVPEHPKQKQITISHNKLPVRQCOQSPN 540
DB 481 MSSPNGSNVNRASPVNASVONWNSVSSVPEHPKQKQITISHNKLPVRQCOQSPN 540
QY 541 LHSNLSLENTKPVPSSTIITNSAVQSTNSASTMSVSSKVTKPIPRSSCSQPVNMGSKLN 600
DB 541 LHSNLSLENTKPVPSSTIITNSAVQSTNSASTMSVSSKVTKPIPRSSCSQPVNMGSKLN 600
QY 601 SSVLPVYGAESESDSDEESKGLGKENGIGTIVSSHSPGQDADEEATPHELOPMTLNGA 660
DB 601 SSVLPVYGAESESDSDEESKGLGKENGIGTIVSSHSPGQDADEEATPHELOPMTLNGA 660
QY 661 NSADSDSPKENGGLAPDGASCOQPALHSENENPAKANGLPGLMPAPLLSLPDKKILETF 720
DB 661 NSADSDSPKENGGLAPDGASCOQPALHSENENPAKANGLPGLMPAPLLSLPDKKILETF 720
QY 721 RLSNKLGSTDEMSAPGAERGGPDRDAPQPGSPAAESLEEDAAA 767
DB 721 RLSNKLGSTDEMSAPGAERGGPDRDAPQPGSPAAESLEEDAAA 767

RESULT 2
US-10-379-981-9
; Sequence 9, Application US/10379981

; Publication No. US20040001820A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pharmaceuticals, Inc.
; APPLICANT: Hahn, Chang S
; APPLICANT: Liu, Hong S
; TITLE OF INVENTION: HUMAN DEUBIQUITINATING PROTEASE GENE ON CHROMOSOME 7 AND ITS
; FILE OF INVENTION: MURINE ORTHOLOG
; CURRENT APPLICATION NUMBER: US/10/379,981
; PRIORITY FILING DATE: 2003-03-05
; PRIORITY FILING DATE: 2002-09-08
; PRIORITY FILING DATE: 2002-03-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 9
; LENGTH: 1316
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-379-981-9

Query Match 85.9%; Score 666; DB 15; Length 1316;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 766; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTIVDKASSSDPSAYONQPGSSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVPGA 60
DB 1 MTIVDKASSSDPSAYONQPGSSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVPGA 60
QY 61 VVYSSSVDPKSPKQKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
DB 61 VVYSSSVDPKSPKQKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTTPPLANYMLSHSHSKTCHAEFGCMCTMQAHTTQALSNPGDVVKPMFVI 180
DB 121 FANAALQCLTYTTPPLANYMLSHSHSKTCHAEFGCMCTMQAHTTQALSNPGDVVKPMFVI 180
QY 181 NEMRRITARHLRFNGQDAHEFLQYTVDMQKACLNGSNKLDRTHTQATTLVCOIFGGYLR 240
DB 181 NEMRRITARHLRFNGQDAHEFLQYTVDMQKACLNGSNKLDRTHTQATTLVCOIFGGYLR 240
QY 241 RVKCLNCKGVSTDFDYLITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSCCKKMWPA 300
DB 241 RVKCLNCKGVSTDFDYLITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSCCKKMWPA 300
QY 301 SKRFTTHRSSNVLTSLKRFANFTGKIAKDVKYPEYLDIRPYMSOPNGEPIVYVLYAVL 360
DB 301 SKRFTTHRSSNVLTSLKRFANFTGKIAKDVKYPEYLDIRPYMSOPNGEPIVYVLYAVL 360
QY 361 VHTGFNCHAGHYFCYIKASNGLWYQNDSTVSTDIRSVLSQAYVLYFIRSHDVKNNGE 420
DB 361 VHTGFNCHAGHYFCYIKASNGLWYQNDSTVSTDIRSVLSQAYVLYFIRSHDVKNNGE 420
QY 421 LTHPTSPGQSSPRPVISQVVTNKQAAFGFTGQPLPSHMIKNPPLHNGTGPLKDTFSSS 480
DB 421 LTHPTSPGQSSPRPVISQVVTNKQAAFGFTGQPLPSHMIKNPPLHNGTGPLKDTFSSS 480
QY 481 MSSPNGSNVNRASPVNASVONWNSVSSVPEHPKQKQITISHNKLPVRQCOQSPN 540
DB 481 MSSPNGSNVNRASPVNASVONWNSVSSVPEHPKQKQITISHNKLPVRQCOQSPN 540
QY 541 LHSNLSLENTKPVPSSTIITNSAVQSTNSASTMSVSSKVTKPIPRSSCSQPVNMGSKLN 600
DB 541 LHSNLSLENTKPVPSSTIITNSAVQSTNSASTMSVSSKVTKPIPRSSCSQPVNMGSKLN 600
QY 601 SSVLPVYGAESESDSDEESKGLGKENGIGTIVSSHSPGQDADEEATPHELOPMTLNGA 660
DB 601 SSVLPVYGAESESDSDEESKGLGKENGIGTIVSSHSPGQDADEEATPHELOPMTLNGA 660
QY 661 NSADSDSPKENGGLAPDGASCOQPALHSENENPAKANGLPGLMPAPLLSLPDKKILETF 720
DB 661 NSADSDSPKENGGLAPDGASCOQPALHSENENPAKANGLPGLMPAPLLSLPDKKILETF 720

QY 721 RLSNKLKSGTDEMSAPGAERGGPPEDRDAEPQGSPPAAESLEEDPAAA 767
Db 721 RLSNKLKSGTDEMSAPGAERGGPPEDRDAEPQGSPPAAESLEEDPAAA 767

RESULT 3
US-10-163-547-1
; Sequence 1, Application US/10163547
; Publication No. US20030037350A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; APPLICANT: Myers, Rachel
; APPLICANT: Rudolf-Owen, Laura A.
; TITLE OF INVENTION: No. US20030037350A1el nucleic acid sequences encoding a
; TITLE OF INVENTION: human ubiquitin protease, lipase, dynamin, short chain
; TITLE OF INVENTION: dehydrogenase, and ADAM-TS Metalloprotease and uses
; TITLE OF INVENTION: therefore
; FILE REFERENCE: 35800/247836(5900-243)
; CURRENT APPLICATION NUMBER: US/10/163,547
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 09/407,356
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 09/704,918
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: 09/435,311
; PRIOR FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: 09/796,100
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/185,503
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: 09/781,598
; PRIOR FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/182,009
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: 09/782,952
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,408
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 09/495,005
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 762
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-163-547-1
Query Match 85.3%; Score 661; DB 14; Length 762;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 761; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MTIVDKASESDPSAYQNPQGSSEAVSGDMDAGSAGVSSINDVSNHTLSLGPVGA 60
Db 1 MTIVDKASESDPSAYQNPQGSSEAVSGDMDAGSAGVSSINDVSNHTLSLGPVGA 60
QY 61 VVYSSSVDPDKSPQKQDQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
Db 61 VVYSSSVDPDKSPQKQDQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTPPLANYMLSHESKTCCHAEFCMWCCTMQAHITQALSNPGDVIKPMFVI 180
Db 121 FANAALQCLTYTPPLANYMLSHESKTCCHAEFCMWCCTMQAHITQALSNPGDVIKPMFVI 180
QY 181 NEMRRIARHLRFGNQEDAEHFLQYTVDMQKACLNGSKNLDHRTQATTLVCOIFGGYLRS 240
Db 181 NEMRRIARHLRFGNQEDAEHFLQYTVDMQKACLNGSKNLDHRTQATTLVCOIFGGYLRS 240
QY 241 RVKCLNCKGVSTDFPYLDITLLEIKAAQSVNKALQFVKPEOLDGNSYKCSKCKMMPA 300
Db 241 RVKCLNCKGVSTDFPYLDITLLEIKAAQSVNKALQFVKPEOLDGNSYKCSKCKMMPA 300

QY 301 SKRFTTHRSSNVLTLSIKRFANFTGGKIAKDVKYPEYLDIRPYMSQNGREPIVVVLAVL 360
Db 301 SKRFTTHRSSNVLTLSIKRFANFTGGKIAKDVKYPEYLDIRPYMSQNGREPIVVVLAVL 360
QY 361 VHTGFENCHAGHYFCYIKASNGLWYQMNDSIVSTDIRSVLSQQAYVLFYIRSHDVKNNGE 420
Db 361 VHTGFENCHAGHYFCYIKASNGLWYQMNDSIVSTDIRSVLSQQAYVLFYIRSHDVKNNGE 420
QY 421 LTHPTSPGQSSPRPVISQRVVTNKQAAPGFIGQLPSHMIKNPPHUNGTCPLKDTFSSS 480
Db 421 LTHPTSPGQSSPRPVISQRVVTNKQAAPGFIGQLPSHMIKNPPHUNGTCPLKDTFSSS 480
QY 481 MSSPENGSSVNRASPVNASASQVQNSVNRSSVPEHPKKOKITITSIHNKLPVROCQOPN 540
Db 481 MSSPENGSSVNRASPVNASASQVQNSVNRSSVPEHPKKOKITITSIHNKLPVROCQOPN 540
QY 541 LHSNLENPTKVPSSITITNSAVQSTSNASTMSVSSKVTPIPRSESCSQPVNMGSKLN 600
Db 541 LHSNLENPTKVPSSITITNSAVQSTSNASTMSVSSKVTPIPRSESCSQPVNMGSKLN 600
QY 601 SSVLPVYGAESSEDSDESKGLKENGIGITIVSSHSGQDAEDEATPHELQEPMTLNGA 660
Db 601 SSVLPVYGAESSEDSDESKGLKENGIGITIVSSHSGQDAEDEATPHELQEPMTLNGA 660
QY 661 NSADSDSDPKENGLAPDQASCOGAPALHSENPFAKANGLPGLKMPAPLLSLPEDKILETF 720
Db 661 NSADSDSDPKENGLAPDQASCOGAPALHSENPFAKANGLPGLKMPAPLLSLPEDKILETF 720
QY 721 RLSNKLKSGTDEMSAPGAERGGPPEDRDAEPQGSPPAAESLEE 762
Db 721 RLSNKLKSGTDEMSAPGAERGGPPEDRDAEPQGSPPAAESLEE 762

RESULT 4
US-09-864-761-47587
; Sequence 47587, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-29
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 47587
 ; LENGTH: 112
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AC004895.2
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.72
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
 ; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.76
 ; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.84
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.66
 ; OTHER INFORMATION: SWISSPROT HIT: P51612, EVALUE 4.00e+00
 ; OTHER INFORMATION: EST_HUMAN HIT: AW575816.1, EVALUE 3.00e-54
 US-09-864-761-47587
 Query Match 14.5%; Score 112; DB 9; Length 112;
 Best Local Similarity 100.0%; Pred. No. 2.3e-100;
 Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 591 PVNMGSKLNSVLPVYGAESSDSEESKGLGKENGIGTIVSSHPGQDADEEATPHE 650
 DB 1 PVNMGSKLNSVLPVYGAESSDSEESKGLGKENGIGTIVSSHPGQDADEEATPHE 60
 QY 651 LQBPMTLNGANSADSDPKENGAPDGASCQCPALHSENPFKANGLPK 702
 DB 61 LQBPMTLNGANSADSDPKENGAPDGASCQCPALHSENPFKANGLPK 112
 RESULT 5
 US-10-379-981-12
 ; Sequence 12, Application US/10379981
 ; Publication No. US20040001820A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Aventis Pharmaceuticals, Inc.
 ; APPLICANT: Hahn, Chang S
 ; APPLICANT: Liu, Hong S
 ; TITLE OF INVENTION: HUMAN DEUBIQUITINATING PROTEASE GENE ON CHROMOSOME 7 AND ITS
 ; FILE REFERENCE: US/2002-0022 WO PCT
 ; CURRENT APPLICATION NUMBER: US/10/379,981
 ; PRIOR FILING DATE: 2003-03-05
 ; PRIOR APPLICATION NUMBER: GB 0218518.9
 ; PRIOR FILING DATE: 2002-09-08
 ; PRIOR APPLICATION NUMBER: US 60/366,601
 ; PRIOR FILING DATE: 2002-03-22
 ; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 12
 ; LENGTH: 1326
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-379-981-12
 Query Match 10.3%; Score 80; DB 15; Length 1326;
 Best Local Similarity 100.0%; Pred. No. 3.1e-68;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 191 RFNQEDAEHFLQYTVDMQKACLNGSNKLDHHTQATTLLVCQIFGGYLRVRKCLNCKGV 250

Query Match 2.1%; Score 16; DB 12; Length 248;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 16; Conservative 0; Mismatches 0; Indels

; PRIOR FILING DATE: 2000-08-04

Query Match 2.1%; Score 16; DB 12; Length 248;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 16; Conservative 0; Mismatches 0; Indels

RESULT 14
US-10-408-765A-2752
; Sequence 2752, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale B.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 66088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2752
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2752

Query Match 2.1%; Score 16; DB 16; Length 472;
Best Local Similarity 100.0%; Pred. No. 2.3e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LOCLTYTPPLANYMLS 141
|||
DB 95 LOCLTYTPPLANYMLS 110
|||

RESULT 15
US-09-372-348-11
; Sequence 11, Application US/09372348A
; Publication No. US2003002805A1
; GENERAL INFORMATION:
; APPLICANT: Hedrick, Joseph A.
; APPLICANT: Wang, Luquan
; APPLICANT: Zlotnik, Albert
; APPLICANT: Murgolo, Nicholas J.
; APPLICANT: Greene, Jonathan R.
; APPLICANT: Johnston, James A.
; APPLICANT: Bazal, J. Fernando
; TITLE OF INVENTION: Mammalian Proteins; Related Reagents and Methods
; FILE REFERENCE: DX0884K(2d)
; CURRENT APPLICATION NUMBER: US/09/372,348A
; CURRENT FILING DATE: 1999-08-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 530
; TYPE: PRT
; ORGANISM: primate
US-09-372-348-11

Query Match 2.1%; Score 16; DB 10; Length 530;
Best Local Similarity 100.0%; Pred. No. 2.5e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LOCLTYTPPLANYMLS 141
|||
DB 95 LOCLTYTPPLANYMLS 110
|||

RESULT 16
US-10-371-905A-12
; Sequence 12, Application US/10371905A
; Publication No. US20030224969A1

; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.3
US-10-371-905A-12

Query Match 2.1%; Score 16; DB 12; Length 530;
Best Local Similarity 100.0%; Pred. No. 2.5e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LOCLTYTPPLANYMLS 141
|||
DB 95 LOCLTYTPPLANYMLS 110
|||

RESULT 17
US-10-371-905A-16
; Sequence 16, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.6
US-10-371-905A-16

Query Match 2.1%; Score 16; DB 12; Length 530;
Best Local Similarity 100.0%; Pred. No. 2.5e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LOCLTYTPPLANYMLS 141
|||
DB 95 LOCLTYTPPLANYMLS 110
|||

RESULT 18
US-10-371-905A-18
; Sequence 18, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 530

```
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.7
US-10-371-905A-18

Query Match
Best Local Similarity 100.0%; DB 12; Length 530;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141
Db 95 LQCLTYTPPLANYMLS 110

RESULT 19
US-10-371-905A-30
; Sequence 30, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB8.5
US-10-371-905A-30

Query Match
Best Local Similarity 100.0%; DB 12; Length 530;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141
Db 95 LQCLTYTPPLANYMLS 110

RESULT 20
US-10-371-905A-34
; Sequence 34, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB8.7
US-10-371-905A-34

Query Match
Best Local Similarity 100.0%; DB 12; Length 530;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141
Db 95 LQCLTYTPPLANYMLS 110

RESULT 21
US-10-107-695B-2
; Sequence 2, Application US/10107695B
; Publication No. US20030022201A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 68999, Human Ubiquitin
; TITLE OF INVENTION: Carboxyl-Terminal Hydrolase Family Member and Uses Therefor
; FILE REFERENCE: MPI01-076PIRCPIM
; CURRENT APPLICATION NUMBER: US/10/107,695B
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 60/279,184
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION:
US-10-107-695B-2

Query Match
Best Local Similarity 100.0%; DB 14; Length 530;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141
Db 95 LQCLTYTPPLANYMLS 110

RESULT 22
US-10-371-905A-20
; Sequence 20, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/C185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.8
US-10-371-905A-20

Query Match
Best Local Similarity 100.0%; DB 12; Length 565;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141
Db 130 LQCLTYTPPLANYMLS 145

RESULT 23
US-10-371-905A-14
; Sequence 14, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; OTHER INFORMATION:
; APPLICANT: Liu, Hong
```

```
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.5
US-10-371-905A-14

Query Match      2.1%; Score 16; DB 12; Length 574;
Best Local Similarity 100.0%; Pred. No. 2.7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      126 LQCLTYTPPLANYMLS 141
      |||||||
Db      130 LQCLTYTPPLANYMLS 145

RESULT 24
US-10-371-905A-24
; Sequence 24, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.11
US-10-371-905A-24

Query Match      2.1%; Score 16; DB 12; Length 633;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      126 LQCLTYTPPLANYMLS 141
      |||||||
Db      198 LQCLTYTPPLANYMLS 213

RESULT 25
US-10-371-905A-8
; Sequence 8, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1021
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.2a
US-10-371-905A-8

Query Match      2.1%; Score 16; DB 12; Length 1021;
Best Local Similarity 100.0%; Pred. No. 4.7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      126 LQCLTYTPPLANYMLS 141
      |||||||
Db      11 LQCLTYTPPLANYMLS 26

RESULT 26
US-10-371-905A-28
; Sequence 28, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB8.3
US-10-371-905A-28

Query Match      1.9%; Score 15; DB 12; Length 497;
Best Local Similarity 100.0%; Pred. No. 2.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      127 LQCLTYTPPLANYMLS 141
      |||||||
Db      96 LQCLTYTPPLANYMLS 110

RESULT 27
US-09-372-348-13
; Sequence 13, Application US/09372348A
; Publication No. US20030028005A1
; GENERAL INFORMATION:
; APPLICANT: Hedrick, Joseph A.
; APPLICANT: Wang, Juquan
; APPLICANT: Zlotnik, Albert
; APPLICANT: Murgolo, Nicholas J.
; APPLICANT: Greene, Jonathan R.
; APPLICANT: Johnston, James A.
; APPLICANT: Bazar, J. Fernando
; TITLE OF INVENTION: Mammalian Proteins; Related Reagents and Methods
; FILE REFERENCE: DX0884K(2d)
; CURRENT APPLICATION NUMBER: US/09/372,348A
; CURRENT FILING DATE: 1999-08-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Primate
; OTHER INFORMATION:
US-09-372-348-13

Query Match      1.8%; Score 14; DB 10; Length 245;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      310 SNVLTSLKRFANF 323
```

```
Db          96 SNVLTSLKRPANF 109
|||||
RESULT 28
US-09-372-348-15
; Sequence 15, Application US/09372348A
; Publication No. US20030028005A1
; GENERAL INFORMATION:
; APPLICANT: Hedrick, Joseph A.
; APPLICANT: Wang, Luquan
; APPLICANT: Zlotnik, Albert
; APPLICANT: Murgolo, Nicholas J.
; APPLICANT: Greene, Jonathan R.
; APPLICANT: Johnston, James A.
; TITLE OF INVENTION: Mammalian Proteins; Related Reagents and Methods
; FILE REFERENCE: DX0884K(2d)
; CURRENT APPLICATION NUMBER: US/09/372,348A
; CURRENT FILING DATE: 1999-08-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 360
; TYPE: PRT
; ORGANISM: primate
US-09-372-348-9
Query Match          1.7%; Score 13; DB 10; Length 360;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          398 SVLSQQAYVLFYI 410
          |||||
Db          191 SVLSQQAYVLFYI 203
          |||||

RESULT 31
US-10-371-905A-6
; Sequence 6, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.1b
US-10-371-905A-6
Query Match          1.7%; Score 13; DB 12; Length 530;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          398 SVLSQQAYVLFYI 410
          |||||
Db          361 SVLSQQAYVLFYI 373
          |||||

RESULT 32
US-10-371-905A-26
; Sequence 26, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
; LENGTH: 530
; TYPE: PRT
US-09-372-348-15
; Sequence 15, Application US/09372348A
; Publication No. US20030028005A1
; GENERAL INFORMATION:
; APPLICANT: Hedrick, Joseph A.
; APPLICANT: Wang, Luquan
; APPLICANT: Zlotnik, Albert
; APPLICANT: Murgolo, Nicholas J.
; APPLICANT: Greene, Jonathan R.
; APPLICANT: Johnston, James A.
; TITLE OF INVENTION: Mammalian Proteins; Related Reagents and Methods
; FILE REFERENCE: DX0884K(2d)
; CURRENT APPLICATION NUMBER: US/09/372,348A
; CURRENT FILING DATE: 1999-08-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 748
; TYPE: PRT
; ORGANISM: primate
US-09-372-348-15
Query Match          1.8%; Score 14; DB 10; Length 748;
Best Local Similarity 100.0%; Pred. No. 0.00031;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          310 SNVLTSLKRPANF 323
          |||||
Db          102 SNVLTSLKRPANF 115
          |||||

RESULT 29
US-09-972-525-1
; Sequence 1, Application US/09972525
; Publication No. US20020086407A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 23484, A NOVEL HUMAN UBIQUITIN
; FILE REFERENCE: 5800-51
; CURRENT APPLICATION NUMBER: US/09/972,525
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/408,865
; PRIOR FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1123
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-525-1
Query Match          1.8%; Score 14; DB 12; Length 1123;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          310 SNVLTSLKRPANF 323
          |||||
Db          321 SNVLTSLKRPANF 334
          |||||

RESULT 30
US-09-372-348-9
; Sequence 9, Application US/09372348A
; Publication No. US20030028005A1
; GENERAL INFORMATION:
; APPLICANT: Hedrick, Joseph A.
; APPLICANT: Wang, Luquan
```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB8.1
US-10-371-905A-26

Query Match 1.7%; Score 13; DB 12; Length 530;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 398 SVLSQAYVLFYI 410
Db 361 SVLSQAYVLFYI 373
|||||

RESULT 33

US-10-371-905A-32
; Sequence 32, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB8.6
US-10-371-905A-32

Query Match 1.7%; Score 13; DB 12; Length 530;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 398 SVLSQAYVLFYI 410
Db 361 SVLSQAYVLFYI 373
|||||

RESULT 34

US-10-371-905A-36
; Sequence 36, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB8.8
US-10-371-905A-36

Query Match 1.7%; Score 13; DB 12; Length 530;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 398 SVLSQAYVLFYI 410

Db 361 SVLSQAYVLFYI 373
|||||

RESULT 35

US-10-371-905A-4
; Sequence 4, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.1a
US-10-371-905A-4

Query Match 1.7%; Score 13; DB 12; Length 1016;
Best Local Similarity 100.0%; Pred. No. 0.0038;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 398 SVLSQAYVLFYI 410
Db 847 SVLSQAYVLFYI 859
|||||

RESULT 36

US-10-424-599-243156
; Sequence 243156, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 243156
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_61599C.1.pap
US-10-424-599-243156

Query Match 1.5%; Score 12; DB 12; Length 150;
Best Local Similarity 100.0%; Pred. No. 0.0062;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLAN 137
Db 39 LQCLTYTPPLAN 50
|||||

RESULT 37

US-10-371-905A-10
; Sequence 10, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang

```

; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0195
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.2b
US-10-371-905A-10

```

Query Match 1.4%; Score 11; DB 12; Length 530;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 11: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 400 LSQQAYVLFYI 410
|||
Db 363 LSQQAYVLFYI 373

RESULT 38

```

US-10-424-599-276474
; Sequence 276474, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 276474
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_91676C.1.pep
US-10-424-599-276474

```

```
Query Match      1.4%; Score 11; DB 12; Length 623;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 11: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	111	AGLQNLGNTCF	121
Rb	16	AGLQNLGNTCF	26

RESULT 39

```

US-10-371-905A-22
; Sequence 22, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Lib, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 715
; TYPE: PRT

```

```

; ORGANISM: Homo sapiens
;
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.10
US-10-371-905A-22

```

Query Match 1.4%; Score 11; DB 12; Length 715;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 11: Conservative 0; Mismatches 0; Indels

Qy 400 LSQQAYVLFYI 410
db 548 LSQQAYVLFYI 558

RESULT 40

```

US-10-437-963-132523
; Sequence 132523, Application US/10437963
; Publication No. US20040123343A1
;
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihwei
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
;
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;
; FILE REFERENCE: 38-21(53221)B
;
; CURRENT APPLICATION NUMBER: US/10/437,963
;
; CURRENT FILING DATE: 2003-05-14
;
; NUMBER OF SEQ ID NOS: 204966
;
; SEQ ID NO 132523
;
; LENGTH: 844
;
; TYPE: PRT
;
; ORGANISM: Oryza sativa
;
; FEATURE:
;
; NAME/KEY: unsure
;
; LOCATION: (1)..(844)
;
; OTHER INFORMATION: unsure at all Xaa locations
;
; FEATURE:
;
; OTHER INFORMATION: Clone ID: PAT_MRT4530_34484C.1.pgp
;
; US-10-437-963-132523

```

Query Match 1.4%; Score 11; DB 16; Length 844;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 11: Conservative 0; Mismatches 0; Indels

```

Qy      110 GAGLQNLGNTC 120
          |||||
Db      99 GAGLQNLGNTC 109

```

RESIT.T 41

```

US-10-424-599-246594
; Sequence 246594, Application US/10424599
; Publication No. US20040031072A1
;
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21 (53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 246594
; LENGTH: 240
; TYPE: PRT

```


; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(240)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_64705C.1.pep
US-10-424-599-246594

Query Match 1.3%; Score 10; DB 12; Length 240;
Best Local Similarity 100.0%; Pred. No. 0.85;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 GLNLGNTCF 121
|||
Db 129 GLNLGNTCF 138
|||

RESULT 42

US-10-371-905A-38
; Sequence 38, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB8.11
US-10-371-905A-38

Query Match 1.3%; Score 10; DB 12; Length 417;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 VYLYAVLVH 362
|||
Db 224 VYLYAVLVH 233
|||

RESULT 43

US-10-424-599-223768
; Sequence 223768, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 223768
; LENGTH: 549
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_44092C.1.pep
US-10-424-599-223768

Query Match 1.3%; Score 10; DB 12; Length 549;

Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 193 GNQEDAHEFL 202
|||
Db 110 GNQEDAHEFL 119
|||

RESULT 44

US-10-425-114-47268
; Sequence 47268, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 47268
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700856585_FLI.pep
US-10-425-114-47268

Query Match 1.3%; Score 10; DB 12; Length 1003;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 GNQEDAHEFL 202
|||
Db 565 GNQEDAHEFL 574
|||

Search completed: August 10, 2004, 20:22:35
Job time : 52 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 20:14:12 ; Search time 21 Seconds
(without alignments)
1905.244 Million cell updates/sec

Title: US-10-049-745-4
Perfect score: 775
Sequence: 1 MTIVDKASSSDPSAYQNQP.....AASLEPDAASLFPFSEG 775

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 10

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents AA.*
1: /cgm2_6/ptodata/2/iaa/5A COMB.pcp.*
2: /cgm2_6/ptodata/2/iaa/5B COMB.pcp.*
3: /cgm2_6/ptodata/2/iaa/6A COMB.pcp.*
4: /cgm2_6/ptodata/2/iaa/6B COMB.pcp.*
5: /cgm2_6/ptodata/2/iaa/PCTUS COMB.pcp.*
6: /cgm2_6/ptodata/2/iaa/backfiles.pcp.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	1.8	1123	4	US-09-408-865-1
2	13	1.7	494	3	US-09-019-095A-24

ALIGNMENTS

RESULT 1
US-09-408-865-1
; Sequence 1, Application US/09408865A
; Patent No. 6329171
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 23484, A NOVEL HUMAN UBIQUITIN
; FILE REFERENCE: 5800-51
; CURRENT APPLICATION NUMBER: US/09/408,865A
; CURRENT FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1123
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-408-865-1

Query Match 1.8%; Score 14; DB 4; Length 1123;
Best Local Similarity 100.0%; Pred. No. 0.00012;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 310 SNVLTSLKRFANF 323
| | | | | | | | | |
Db 321 SNVLTSLKRFANF 334

RESULT 2
US-09-019-095A-24
; Sequence 24, Application US/09019095A
; Patent No. 6287858
; GENERAL INFORMATION:
; APPLICANT: D'Andrea, Alan D.
; APPLICANT: Zhu, Yuan
; TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
; FILE REFERENCE: DFCT-435p2A2
; CURRENT APPLICATION NUMBER: US/09/019,095A
; CURRENT FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US96/12884
; PRIOR FILING DATE: 1996-08-07
; PRIOR APPLICATION NUMBER: US 60/002,066
; PRIOR FILING DATE: 1995-08-09
; PRIOR APPLICATION NUMBER: US 60/019,787
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 494
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-019-095A-24

Query Match 1.7%; Score 13; DB 3; Length 494;
Best Local Similarity 100.0%; Pred. No. 0.00052;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 LTYTPPLANYMLS 141
| | | | | | | | | |
Db 63 LTYTPPLANYMLS 75

Search completed: August 10, 2004, 20:17:40
Job time : 21 secs

THIS PAGE BLANK (USPTO)

QY 1921 GAGAGTCAAGGGGCTGGGCAAGAGAAATGGGATTGTGATGAGCTCCACTCT 1980
Db 2003 GAGAGTCAAGGGGCTGGGCAAGAGAAATGGGATTGTGATGAGCTCCACTCT 2062
QY 1981 CCCGGCAAGATGCCGAAGATGAGGAGGCCACTCCGCAAGAGCTTCAAGAACCCATGACC 2040
Db 2063 CCCGGCAAGATGCCGAAGATGAGGAGGCCACTCCGCAAGAGCTTCAAGAACCCATGACC 2122
QY 2041 CTAAACGGTGTCTAATGATGCGAGACAGACAGTACCCGAAAGAAACCGGCTAGCGCT 2100
Db 2123 CTAAACGGTGTCTAATGATGCGAGACAGACAGTACCCGAAAGAAACCGGCTAGCGCT 2182
QY 2101 GATGGTGCACCTGCCAGGCCAGCTCCCTGCACTCAGAAATCCCTTTGCTTAAGCA 2160
Db 2183 GATGGTGCACCTGCCAGGCCAGCTCCCTGCACTCAGAAATCCCTTTGCTTAAGCA 2242
QY 2161 AACGGTCTTCTCGAAAGCTTGATGCTGCTCTCTTTGCTGTCTCTCCAGAAAGCAAAATC 2220
Db 2243 AACGGTCTTCTCGAAAGCTTGATGCTGCTCTCTTTGCTGTCTCTCCAGAAAGCAAAATC 2302
QY 2221 TTAGAGACCTTCAGGCTTAGCAACAACTGAAGGCTCGACGGATGAAATGAGTGCACCT 2280
Db 2303 TTAGAGACCTTCAGGCTTAGCAACAACTGAAGGCTCGACGGATGAAATGAGTGCACCT 2362
QY 2281 GGAGCAGAGAGGGGCTCCCGAGGACCGGACGCGGAGCCCTCAGCTGGCAGCCCGCC 2340
Db 2363 GGAGCAGAGAGGGGCTCCCGAGGACCGGACGCGGAGCCCTCAGCTGGCAGCCCGCC 2422
QY 2341 GCCGAATCCCTGGAGGACCGAGATGCGGCCGCAAGCTT 2378
Db 2423 GCCGAATCCCTGGAGGACCGAGATGCGGCCGCGGCT 2460

RESULT 2

US-10-163-547-2
; Sequence 2, Application US/10163547
; Publication No. US20030037350A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksman, Maria A.
; APPLICANT: Myers, Rachel
; APPLICANT: Rudolf-Owen, Laura A.
; TITLE OF INVENTION: No. US20030037350A1el nucleic acid sequences encoding a
; TITLE OF INVENTION: human ubiquitin protease, lipase, dynamin, short chain
; TITLE OF INVENTION: dehydrogenase, and ADAM-15 Metalloprotease and uses
; TITLE OF INVENTION: therefore
; FILE REFERENCE: 35800/247836(5800-243)
; CURRENT APPLICATION NUMBER: US/10/163,547
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 09/407,356
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 09/704,918
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: 09/435,311
; PRIOR FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: 09/796,100
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/185,503
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: 09/781,598
; PRIOR FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/182,009
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: 09/782,952
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,408
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 09/496,005
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2

LENGTH: 2347
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (0)..(0)
OTHER INFORMATION: 23431 Ubiquitin protease
FEATURE:
NAME/KEY: CDS
LOCATION: (61)...(2347)
US-10-163-547-2

Query Match 97.0%; Score 2339.2; DB 15; Length 2347;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2341; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 16 GCGGCGGCGGCGGCGGAGGGGGATGGAGCGCGCGGAGCGGGGTGAGAGTTGAACAATG 75
Db 4 GCGTCCGCGCGCGGAGGGGGATGGAGCGCGCGGAGCGGGGTGAGAGTTGAACAATG 63
QY 76 ACCATAGTTGAACAAGCTTCTGAATCTTTCAGACCATCAGCTATCAGAAATCAGCTGCG 135
Db 64 ACCATAGTTGAACAAGCTTCTGAATCTTTCAGACCATCAGCTATCAGAAATCAGCTGCG 123
QY 136 AGCTCCGAGGCGAGTCTACCTGGGAGACATGGATGAGGCTTCTGCCAGCTGGGGTCTGTG 195
Db 124 AGCTCCGAGGCGAGTCTACCTGGGAGACATGGATGAGGCTTCTGCCAGCTGGGGTCTGTG 183
QY 196 TCTTCATTGAATGATGTGTCAAAATCACACATCTTCTTTAGGACCATGACCTGGTGTCTGA 255
Db 184 TCTTCATTGAATGATGTGTCAAAATCACACATCTTCTTTAGGACCATGACCTGGTGTCTGA 243
QY 256 GTTTATTGCGATTGATCTGTATACCTGATTAATCAAAACCATCAACCAAAAGGATCAAGCC 315
Db 244 GTTTATTGCGATTGATCTGTATACCTGATTAATCAAAACCATCAACCAAAAGGATCAAGCC 303
QY 316 CTAGTGATGCGATCGCTCCACAGAAAGTCTTTTCCCATCTGAGAAATTTGTCTT 375
Db 304 CTAGTGATGCGATCGCTCCACAGAAAGTCTTTTCCCATCTGAGAAATTTGTCTT 363
QY 376 AAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTCTT 435
Db 364 AAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTCTT 423
QY 436 GCGAATGAGCACTGAGCTGTTTAACTTACACACACCTCTTGCCAAATACATGCTATCA 495
Db 424 GCGAATGAGCACTGAGCTGTTTAACTTACACACACCTCTTGCCAAATACATGCTATCA 483
QY 496 CATGAACACTCCAAACATGTCATGCAAGGCTTTTGTATGATCTGTACAAATGCAAGCA 555
Db 484 CATGAACACTCCAAACATGTCATGCAAGGCTTTTGTATGATCTGTACAAATGCAAGCA 543
QY 556 CATATTACCCAGGCACTCAGTAATCTGGGAGCGTTATTAAACCAATGTTTGTCTCAAT 615
Db 544 CATATTACCCAGGCACTCAGTAATCTGGGAGCGTTATTAAACCAATGTTTGTCTCAAT 603
QY 616 GAGATCGCGGTATAGCTAGGACCTCCGTTTGGAAACCAAGAGATGCCCATGAATTC 675
Db 604 GAGATCGCGGTATAGCTAGGACCTCCGTTTGGAAACCAAGAGATGCCCATGAATTC 663
QY 676 CTTCAATACACTGTTGATGCTATGCAAAAGCATGCTTGAATGGCAGCAATAAATTAGAC 735
Db 664 CTTCAATACACTGTTGATGCTATGCAAAAGCATGCTTGAATGGCAGCAATAAATTAGAC 723
QY 736 AGACACACCCAGGCGCACCTCTTGTGTCAGATATTTGGAGGATACCTAAGATCTAGA 795
Db 724 AGACACACCCAGGCGCACCTCTTGTGTCAGATATTTGGAGGATACCTAAGATCTAGA 783
QY 796 GTCAAAATGTTAAATGCAAGGGGCTTTCAGATACCTTTTGTATCCATATCTTGTATATA 855
Db 784 GTCAAAATGTTAAATGCAAGGGGCTTTCAGATACCTTTTGTATCCATATCTTGTATATA 843
QY 856 TTGGAGATAAAGGCTGCTCAGAGTGTCAACAGGCAATTTGGAGCAATTTGTGAAGCCGAA 915

Db 844 TTGGAGATAAAGCTGCTCAGAGTGTCACAAAGCAATGGAGCAGTTTGTGAAGCGGAA 903
QY 916 CAGCTTTGATGGAGAAAACCTCGTCAAGTGCAGCAAGTGTAAAAGATGGTTCCAGCTTCA 975
Db 904 CAGCTTTGATGGAGAAAACCTCGTCAAGTGCAGCAAGTGTAAAAGATGGTTCCAGCTTCA 963
QY 976 AAGAGGTTCACTATCCATAGATCCTCTAATGTGTTCTTACACTTTCTCTGAACGTTTGGCA 1035
Db 964 AAGAGGTTCACTATCCATAGATCCTCTAATGTGTTCTTACACTTTCTCTGAACGTTTGGCA 1023
QY 1036 AATTTTACCGGTGGAATAAATTTGCTAAGATGTGAATATACCCTCAGTATCTTTGATATTCGG 1095
Db 1024 AATTTTACCGGTGGAATAAATTTGCTAAGATGTGAATATACCCTCAGTATCTTTGATATTCGG 1083
QY 1096 CCATATATGTTCTCAACCCAAAGGAGGCCAAATTTGTCTACGTTCTTTGATGCAAGTCTGGTC 1155
Db 1084 CCATATATGTTCTCAACCCAAAGGAGGCCAAATTTGTCTACGTTCTTTGATGCAAGTCTGGTC 1143
QY 1156 CACACTGGTTTAAATGGCCATGCTGGCCATTACTCTGCTACATATAAAGCTACGAATGGC 1215
Db 1144 CACACTGGTTTAAATGGCCATGCTGGCCATTACTCTGCTACATATAAAGCTACGAATGGC 1203
QY 1216 CTCCTGGTATCAAAATGAATGACTCCATTTGTATCTACCAAGTATATAGATCGGTACTCAGC 1275
Db 1204 CTCCTGGTATCAAAATGAATGACTCCATTTGTATCTACCAAGTATATAGATCGGTACTCAGC 1263
QY 1276 CAACAAGCCCTATGTGCTCTTTTATATACAGTCCCATGATGTGAAAAATGAGGTGAACCTT 1335
Db 1264 CAACAAGCCCTATGTGCTCTTTTATATACAGTCCCATGATGTGAAAAATGAGGTGAACCTT 1323
QY 1336 ACTCATCCCAACCATAGCCCGGCCAGTCTCTCCCGCCCGCTCATCTACGTCAGCGGTT 1395
Db 1324 ACTCATCCCAACCATAGCCCGGCCAGTCTCTCCCGCCCGCTCATCTACGTCAGCGGTT 1383
QY 1396 GTCAACCAAAACAGGCTGCGCCAGGCTTTATCGGACCAAGTCTCCCTCTCACATGATA 1455
Db 1384 GTCAACCAAAACAGGCTGCGCCAGGCTTTATCGGACCAAGTCTCCCTCTCACATGATA 1443
QY 1456 AAGAAATCCAGCTCACTTAAATGGGACTGGACCAATTTGAAGACAGCGCAAGCTTCATG 1515
Db 1444 AAGAAATCCAGCTCACTTAAATGGGACTGGACCAATTTGAAGACAGCGCAAGCTTCATG 1503
QY 1516 TCAGGTCCTAAACGGGAATTCAGTGTCACAGAGCTAGTCTCTGTTAATGCTTCAGGTTCT 1575
Db 1504 TCAGGTCCTAAACGGGAATTCAGTGTCACAGAGCTAGTCTCTGTTAATGCTTCAGGTTCT 1563
QY 1576 GTCAAAACACTGGTCAGTTAATAGTCTCTCAGTGTATCCAGAACATCTTAAGAAACAAAA 1635
Db 1564 GTCAAAACACTGGTCAGTTAATAGTCTCTCAGTGTATCCAGAACATCTTAAGAAACAAAA 1623
QY 1636 ATTACAAATCAGTATTCACAAAGATTGCTGTTGCGCCAGTGTCAACCTAACCTT 1695
Db 1624 ATTACAAATCAGTATTCACAAAGATTGCTGTTGCGCCAGTGTCAACCTAACCTT 1683
QY 1696 CATAGTAATTTCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCATTAACAATTC 1755
Db 1684 CATAGTAATTTCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCATTAACAATTC 1743
QY 1756 GCAGTACAGTCTACCTCGAAACGATCTACGATGTCAAGTCTTCTAGTAAAGTAAACAAACCG 1815
Db 1744 GCAGTACAGTCTACCTCGAAACGATCTACGATGTCAAGTCTTCTAGTAAAGTAAACAAACCG 1803
QY 1816 ATCCCCCGAGTGAATCCTGCTCCAGCCGCTGATGAATGGCAATCCAGCTGAACCTC 1875
Db 1804 ATCCCCCGAGTGAATCCTGCTCCAGCCGCTGATGAATGGCAATCCAGCTGAACCTC 1863
QY 1876 AGCGTGTGCTGCTGATGCGCGGAGTCTCTGAGGACTCTGACGAGGAGTCAAAAGGGG 1935
Db 1864 AGCGTGTGCTGCTGATGCGCGGAGTCTCTGAGGACTCTGACGNGGAGTCAAAAGGGG 1923
QY 1936 CTGGGCAAGGAGAAATGGGATGGTACGATTTGAGCTCCCACTCTCCCGGCCCAAGATGCC 1995

Db 1924 CTGGCAAGGAGAAATGGGATTGGTACGATTGTGAGCTCCCACTCTCCCGGCCAAGATGCC 1983
QY 1996 GAAGATGAGAGAGGCCACTTCGGCAGAGCTTTCAAGAAACCCATGACCTTAAACGGTGTCTAAT 2055
Db 1984 GAAGATGAGAGAGGCCACTTCGGCAGAGCTTTCAAGAAACCCATGACCTTAAACGGTGTCTAAT 2043
QY 2056 AGTGCAGACAGCCACAGTGTGACCCGAAAGAAACGGCTAGCGCTGTATGTCGCCAGCTGC 2115
Db 2044 AGTGCAGACAGCCACAGTGTGACCCGAAAGAAACGGCTAGCGCTGTATGTCGCCAGCTGC 2103
QY 2116 CAAGGCCAGCTCCCTGCACACTCAGAAAAATCCCTTTGCTAAGGCAAAACGGTCTTCCCTGGA 2175
Db 2104 CAAGGCCAGCTCCCTGCACACTCAGAAAAATCCCTTTGCTAAGGCAAAACGGTCTTCCCTGGA 2163
QY 2176 AAGTTGATGCTCTCTCTTTGCTGTCTCTCCAGAAAGCAAAAATCTTTAGAGACCTTTCAGG 2235
Db 2164 AAGTTGATGCTCTCTCTTTGCTGTCTCTCCAGAAAGCAAAAATCTTTAGAGACCTTTCAGG 2223
QY 2236 CTTAGCAACAAACTGAAAGGCTCGACGGATGAAATGAGTGCACCTGGAGCAGAGAGGGGC 2295
Db 2224 CTTAGCAACAAACTGAAAGGCTCGACGGATGAAATGAGTGCACCTGGAGCAGAGAGGGGC 2283
QY 2296 CTTCCGAGACCGCGACGCGAGCTCAGCTGGCAGCCCCCGCCGCAATCCCTGGAG 2355
Db 2284 CTTCCGAGACCGCGACGCGAGCTCAGCTGGCAGCCCCCGCCGCAATCCCTGGAG 2343
QY 2356 GAGC 2359
Db 2344 GAGC 2347

RESULT 3
US-10-302-172-390
; Sequence 390, Application US/10302172
; Publication No. US20040053250A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: No. US20040053250A1 Arginine-rich Protein-like Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/10/302,172
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/225,251
; PRIOR FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: PCT US02/05095
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 09/799,451
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 950
; SOFTWARE: pt_Fl_genes version 2.0
; SEQ ID NO 390
; LENGTH: 4483
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (191)..(4000)
US-10-302-172-390

Query Match 95.8%; Score 2310; DB 13; Length 4483;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2313; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 61 CAGAGTTGAACAAATACCAATAGTTGACAAAGCTTCTGAATCTTTACAGCCCATCAGCCTAT 120
Db 179 CAGAGTTGAACAAATACCAATAGTTGACAAAGCTTCTGAATCTTTACAGCCCATCAGCCTAT 238
QY 121 CAGAAATCAGCCTGGCAGCTCCGAGGAGTCTACCTGGAGACATGGATGAGGTTCTTCC 180
Db 239 CAGAAATCAGCCTGGCAGCTCCGAGGAGTCTACCTGGAGACATGGATGAGGTTCTTCC 298

QY	181	AGCTGGGGTCTGTGCTTCTTCATTGAATGATGTGTCAAATCACACACTTCTTTTAGACCA	240
Db	299	AGCTGGGGTCTGTGCTTCTTCATTGAATGATGTGTCAAATCACACACTTCTTTTAGACCA	358
QY	241	GTACCTGGTCTGTAGTTTATTTCGAGTTTCACTGTACCTGATAAATCAAAACCATCACCA	300
Db	359	GTACCTGGTCTGTAGTTTATTTCGAGTTTCACTGTACCTGATAAATCAAAACCATCACCA	418
QY	301	CAAAAGGATCAAGCCCTTAGTGATGGCATCGCTCCCTCCACAGAAAGTCTCTTTTCCCATCT	360
Db	419	CAAAAGGATCAAGCCCTTAGTGATGGCATCGCTCCCTCCACAGAAAGTCTCTTTTCCCATCT	478
QY	361	GAGAAGATTGTTCTTTAAGTGGCAACAAATCTCATAGAGTTGGAGCTGGGCTCCAGAATTGG	420
Db	479	GAGAAGATTGTTCTTTAAGTGGCAACAAATCTCATAGAGTTGGAGCTGGGCTCCAGAATTGG	538
QY	421	GGCAATACCTGTTTTGCCAAATGAGCACTGCAGTGTTTAACTACACACACACTCTTGCC	480
Db	539	GGCAATACCTGTTTTGCCAAATGAGCACTGCAGTGTTTAACTACACACACACTCTTGCC	598
QY	481	AAATTACATGCTATCACATGAACACTCCAAACCATGTGCATGAGAAGGCTTTTGTATGATG	540
Db	599	AAATTACATGCTATCACATGAACACTCCAAACCATGTGCATGAGAAGGCTTTTGTATGATG	658
QY	541	TGTACAATGAAGCACAATATTACCAGGCACTCAGTAAATCCTGGGAGGTTATTAAACCA	600
Db	659	TGTACAATGAAGCACAATATTACCAGGCACTCAGTAAATCCTGGGAGGTTATTAAACCA	718
QY	601	ATGTTTTGTCATCAATGAGATGCGCGGTATAGCTAGGCACCTCCGTTTTGGAACCAAGAA	660
Db	719	ATGTTTTGTCATCAATGAGATGCGCGGTATAGCTAGGCACCTCCGTTTTGGAACCAAGAA	778
QY	661	GATGCCCATGAAATTCCTTTCAAATCACCTGTGTGATGCTATGCAGAAAGCATGCTTTGAATGCC	720
Db	779	GATGCCCATGAAATTCCTTTCAAATCACCTGTGTGATGCTATGCAGAAAGCATGCTTTGAATGCC	838
QY	721	AGCAATAAATTAGACAGACACACCGGCGACCACTCTTGTGTCAGATATTGAGGGA	780
Db	839	AGCAATAAATTAGACAGACACACCGGCGACCACTCTTGTGTCAGATATTGAGGGA	898
QY	781	TACCTAAGATCTAGAGTCAAAATGTTTAAATGCAAGGCGTTTCAGATACTTTTGATCCA	840
Db	899	TACCTAAGATCTAGAGTCAAAATGTTTAAATGCAAGGCGTTTCAGATACTTTTGATCCA	958
QY	841	TATCTTGATATAACATTTGGAGATAAAGGCTGCTCAGAGTGTCAACAGGCAATTTGGAGCAG	900
Db	959	TATCTTGATATAACATTTGGAGATAAAGGCTGCTCAGAGTGTCAACAGGCAATTTGGAGCAG	1018
QY	901	TTTGTGAAGCGGGAACAGCTTGATGGAAGAAAATCTCGTACAAAGTGCAGCAAGTGTAAAAG	960
Db	1019	TTTGTGAAGCGGGAACAGCTTGATGGAAGAAAATCTCGTACAAAGTGCAGCAAGTGTAAAAG	1078
QY	961	ATGGTTCGAGCTTCAAGAGGTTTCACTATCCATAGATCCCTTAATGTTCTTACACTTCT	1020
Db	1079	ATGGTTCGAGCTTCAAGAGGTTTCACTATCCATAGATCCCTTAATGTTCTTACACTTCT	1138
QY	1021	CTGAAAACGTTTTGCAAAATTTTACCGGTGGAAAAATTCCTAAGGATGTGAAATACCCCTGAG	1080
Db	1139	CTGAAAACGTTTTGCAAAATTTTACCGGTGGAAAAATTCCTAAGGATGTGAAATACCCCTGAG	1198
QY	1081	TATCTTGATATTTCGGCCATATATGCTCTCAACCCAAACGGAGAGCAATGTGCTACGCTTG	1140
Db	1199	TATCTTGATATTTCGGCCATATATGCTCTCAACCCAAACGGAGAGCAATGTGCTACGCTTG	1258
QY	1141	TATGCAAGTGTGTTCCACACTGGTTTTTAAATGTCATGCTGGCCATTTACTTCTGCTACATA	1200
Db	1259	TATGCAAGTGTGTTCCACACTGGTTTTTAAATGTCATGCTGGCCATTTACTTCTGCTACATA	1318
QY	1201	AAAGCTAGCAATGGCCTCTGGTATCAAAATGAATGACTCCATTGTATCTCACAGTGATATT	1260
Db	1319	AAAGCTAGCAATGGCCTCTGGTATCAAAATGAATGACTCCATTGTATCTCACAGTGATATT	1378
QY	1261	AGATCGGTACTTCAGCAACCAAGCCCTATGTGCTCTTTTTATATCAGGTCCTCATGATGTGAAA	1320

[illegible]

Db 2459 GCCGAATCCCTGGAGGAGCCAGATGCGCGCGGCT 2496

RESULT 4
US-10-379-981-8
; Sequence 8, Application US/10379981
; Publication No. US2004001820A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pharmaceuticals, Inc.
; APPLICANT: Liu, Hong S
; TITLE OF INVENTION: HUMAN DEUBIQUITINATING PROTEASE GENE ON CHROMOSOME 7 AND ITS
; FILE REFERENCE: USAV2002-0022, WO PCT
; CURRENT APPLICATION NUMBER: US/10/379,981
; CURRENT FILING DATE: 2003-03-05
; PRIOR APPLICATION NUMBER: GB 0218518.9
; PRIOR FILING DATE: 2002-09-08
; PRIOR APPLICATION NUMBER: US 60/366,601
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 3951
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-379-981-8

Query Match 95.3%; Score 2299.6; DB 16; Length 3951;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2302; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 73 ATGACCATAGTTCACAAAGCTTCTGATCTTCAGACCCATCAGCCATCAGATCAGCCT 132
Db 1 ATGACCATAGTTCACAAAGCTTCTGATCTTCAGACCCATCAGCCATCAGATCAGCCT 60

QY 133 GCGAGCTCCGAGGAGCTCCACCTGGAGACATGGATGAGCTTCTGCGAGCTGGGGTGT 192
Db 61 GCGAGCTCCGAGGAGCTCCACCTGGAGACATGGATGAGCTTCTGCGAGCTGGGGTGT 120

QY 193 GTGCTTTCATGATGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCG 252
Db 121 GTGCTTTCATGATGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCG 180

QY 253 GTAGTTTATTCGAGTTCATCTGTCACCTGATTAATCAAAACCATCACCACAAAGGATCAA 312
Db 181 GTAGTTTATTCGAGTTCATCTGTCACCTGATTAATCAAAACCATCACCACAAAGGATCAA 240

QY 313 GCGTAGGTGATGGCATCGCTCTCCACAGAAAGTTCCTTTCCCATCTGAGAAAGATTGT 372
Db 241 GCGCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCCTTTCCCATCTGAGAAAGATTGT 300

QY 373 CTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGCAATACCTGT 432
Db 301 CTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGT 360

QY 433 TTTGCCAATGCACACTGACAGTCTTTAACTCACCACACCTCTTGCCTAATTACATGCTA 492
Db 361 TTTGCCAATGCACACTGACAGTCTTTAACTCACCACACCTCTTGCCTAATTACATGCTA 420

QY 493 TCACATGAACACTCCAAAACATGTCATGACAGAGGCTTTTGTATGATGTCACAAATGCAA 552
Db 421 TCACATGAACACTCCAAAACATGTCATGACAGAGGCTTTTGTATGATGTCACAAATGCAA 480

QY 553 GCACATATTACCCAGACACTCAGTATCTCGGGAGCTTTATTAACCAATGTTTGTATC 612
Db 481 GCACATATTACCCAGACACTCAGTATCTCGGGAGCTTTATTAACCAATGTTTGTATC 540

QY 613 AATGAGATCGGGGTATAGTAGGCACTCCGTTTGGAAACCAAGAGATGCCATGAA 672
Db 541 AATGAGATCGGGGTATAGTAGGCACTCCGTTTGGAAACCAAGAGATGCCATGAA 600

QY 673 TTCCTTCAATACACTGTTGATGCTATGCGAAGAGCATGCTTGAATGGCAGCAATAAATTA 732

Db 601 TTCCTTCAATACACTGTTGATGCTATGCGAAGACATGCTTGAATGGCAGCAATAAATTA 660

QY 733 GACAGACACACCCAGCCACCACTCTTTGTTGTCAGATATTTGGAGGATACCTAAGATCT 792

Db 661 GACAGACACACCCAGCCACCACTCTTTGTTGTCAGATATTTGGAGGATACCTAAGATCT 720

QY 793 AGAGTCAATGTTTAAATTCGAAGGGGCTTTCAGATACCTTTGATCCATATCTTTGATATA 852

Db 721 AGAGTCAATGTTTAAATTCGAAGGGGCTTTCAGATACCTTTGATCCATATCTTTGATATA 780

QY 853 ACATTGGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCATTGGAGCAGTTTGTGAAGCCG 912

Db 781 ACATTGGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCATTGGAGCAGTTTGTGAAGCCG 840

QY 913 GAAACAGCTTCATGAGGAGAAAACTCGTACAACTGAGCAAGTGTAAAAAGATGGTTCAGCT 972

Db 841 GAAACAGCTTCATGAGGAGAAAACTCGTACAACTGAGCAAGTGTAAAAAGATGGTTCAGCT 900

QY 973 TCAAGAGGTTCACTATCCATAGATCTCTAATGTTCTTACACTTTCTCTGAAACGTTTT 1032

Db 901 TCAAGAGGTTCACTATCCATAGATCTCTAATGTTCTTACACTTTCTCTGAAACGTTTT 960

QY 1033 GCAAAATTTACCGGTGGAAAAATTTGCTAAGATGTGAAATACCTGAGTATCTTGATATT 1092

Db 961 GCAAAATTTACCGGTGGAAAAATTTGCTAAGATGTGAAATACCTGAGTATCTTGATATT 1020

QY 1093 CGGCCATATATGCTCAACCCAAAGGAGCCAAATTTGCTTACGTTCTTTGATGAGTGTCTG 1152

Db 1021 CGGCCATATATGCTCAACCCAAAGGAGCCAAATTTGCTTACGTTCTTTGATGAGTGTCTG 1080

QY 1153 GTCCACACTGTTTAAATTCGCAATGCTGGCCATTTCTGCTACATATAAAGCTAGCAAT 1212

Db 1081 GTCCACACTGTTTAAATTCGCAATGCTGGCCATTTCTGCTACATATAAAGCTAGCAAT 1140

QY 1213 GGCCTCTGGTATCAAAATGAATGATCTCATTTGATCTACCAAGTATATAGATCGGTACTC 1272

Db 1141 GGCCTCTGGTATCAAAATGAATGATCTCATTTGATCTACCAAGTATATAGATCGGTACTC 1200

QY 1273 AGCCAAACAAGCTATGTCCTTTTATATCAGTCTCCATGATGTGAAAAATGAGGTGAA 1332

Db 1201 AGCCAAACAAGCTATGTCCTTTTATATCAGTCTCCATGATGTGAAAAATGAGGTGAA 1260

QY 1333 CTTTACTCATCCACCCATAGCCCGGCGAGTCTCTCCCGCCCGCTCATCAGTCAGCGG 1392

Db 1261 CTTTACTCATCCACCCATAGCCCGGCGAGTCTCTCTCCCGCCCGCTCATCAGTCAGCGG 1320

QY 1393 GTTGTCAACCAACAAACAGGCTGCGCCAGGCTTTTATCGGACCAAGCTTCCCTCTCACATG 1452

Db 1321 GTTGTCAACCAACAAACAGGCTGCGCCAGGCTTTTATCGGACCAAGCTTCCCTCTCACATG 1380

QY 1453 ATAAAGATTCACCTCATTTAAATGGGACTGGACCAATTTGAAAGACACGCGCAAGCTTCC 1512

Db 1381 ATAAAGATTCACCTCATTTAAATGGGACTGGACCAATTTGAAAGACACGCGCAAGCTTCC 1440

QY 1513 ATGTGAGTCTTAACCGGAAATTCAGTGTCAACAGGCTAGTCTGTTAATGCTTCAGCT 1572

Db 1441 ATGTGAGTCTTAACCGGAAATTCAGTGTCAACAGGCTAGTCTGTTAATGCTTCAGCT 1500

QY 1573 TCTGTCCAAAACCTGGTTCAGTTAATAGGTCCTCAGTATCCAGAACATCTTAAGAAACAA 1632

Db 1501 TCTGTCCAAAACCTGGTTCAGTTAATAGGTCCTCAGTATCCAGAACATCTTAAGAAACAA 1560

QY 1633 AAAATTTACATCAGTATTCACAAAGTTGCCCTGTTGCCAGTGTGAGTCTCAACCTAAC 1692

Db 1561 AAAATTTACATCAGTATTCACAAAGTTGCCCTGTTGCCAGTGTGAGTCTCAACCTAAC 1620

QY 1693 CTTTCAAGTAAATTTCTTTGGAGAACCTTACCAAGCCGTTTCCCTCTTCTACATTAACCAAT 1752

Db 1621 CTTTCAAGTAAATTTCTTTGGAGAACCTTACCAAGCCGTTTCCCTCTTCTACATTAACCAAT 1680

QY 1753 TCTGCAAGTACAGTCTACCTCGAAACGATCTAGAGTGTGAGTCTTCTAGTAAAGTAACAAA 1812

Db	1581	TCTGCAGTACAGTCTACTCGAA	CGCATCTACGATGTGAGTTTCTAGTAAAGTACAAAA	1740
Qy	1813	CCGATCCCCCAGTGAATCTCTGCTCCAGCCGCGTGATGAATGGCAATCCAAGCTGAAC	1872	
Db	1741	CCGATCCCCCAGTGAATCTCTGCTCCAGCCGCGTGATGAATGGCAATCCAAGCTGAAC	1800	
Qy	1873	TCCAGCGTGTGTTGCTGCTATGCGCGCGAGTCCCTCTGAGGACTCTGACGAGAGTCAAG	1932	
Db	1801	TCCAGCGTGTGTTGCTGCTATGCGCGCGAGTCCCTCTGAGGACTCTGACGAGAGTCAAG	1860	
Qy	1933	GGGCTGGCAAGGAGAAATGGATTTGGTACGATTTGTGAGTCTCCCACTCTCCCGGCCAAGAT	1992	
Db	1861	GGGCTGGCAAGGAGAAATGGATTTGGTACGATTTGTGAGTCTCCCACTCTCCCGGCCAAGAT	1920	
Qy	1993	GCCGAAGATGAGGAGCCACTCCCGCACGAGTTCGAAGAACCCATGACCCCTAAACCGGTGCT	2052	
Db	1921	GCCGAAGATGAGGAGCCACTCCCGCACGAGTTCGAAGAACCCATGACCCCTAAACCGGTGCT	1980	
Qy	2053	AATAGTGCAGACGACGAGTGAACCCGGAAGAAACGGCCCTAGCGCCTGATGGTCCGAGC	2112	
Db	1981	AATAGTGCAGACGACGAGTGAACCCGGAAGAAACGGCCCTAGCGCCTGATGGTCCGAGC	2040	
Qy	2113	TGCCAAGCCAGCCTGCCCTGCACCTCAGAAAATCCCTTTGCTAAGGCCAAACCGTCTTCT	2172	
Db	2041	TGCCAAGCCAGCCTGCCCTGCACCTCAGAAAATCCCTTTGCTAAGGCCAAACCGTCTTCT	2100	
Qy	2173	GGAAGTTGATGCTCTGCTCTTTGCTGCTCTCCAGAAAGCAAAAATCTTAGAGACCTTC	2232	
Db	2101	GGAAGTTGATGCTCTGCTCTTTGCTGCTCTCCAGAAAGCAAAAATCTTAGAGACCTTC	2160	
Qy	2233	AGGCTTAGCAACAACCTGAAGGCTCGAGGATGAATGAGTGACCTGGAGCAGAGG	2292	
Db	2161	AGGCTTAGCAACAACCTGAAGGCTCGAGGATGAATGAGTGACCTGGAGCAGAGG	2220	
Qy	2293	GGCCCTCCCGAGGACCGGACCGCGAGCTCAGCCTGGCAGCCCGCCGCGCAATCCCTG	2352	
Db	2221	GGCCCTCCCGAGGACCGGACCGCGAGCTCAGCCTGGCAGCCCGCCGCGCAATCCCTG	2280	
Qy	2353	GAGGAGCCAGATGCGGCGCGCAAGCTT	2378	
Db	2281	GAGGAGCCAGATGCGGCGCGCGCT	2306	
RESULT 5				
US-10-379-981-11				
; Sequence 11, Application US/10379981				
; Publication No. US20040001820A1				
; GENERAL INFORMATION:				
; APPLICANT: Aventis Pharmaceuticals, Inc.				
; APPLICANT: Hahn, Chang S				
; APPLICANT: Liu, Hong S				
; TITLE OF INVENTION: HUMAN DEUBIQUITINATING PROTEASE GENE ON CHROMOSOME 7 AND ITS				
; FILE OF INVENTION: MURINE ORTHOLOG				
; FILE REFERENCE: USAV2002-0022 WO PCT				
; CURRENT APPLICATION NUMBER: US/10/379,981				
; CURRENT FILING DATE: 2003-03-05				
; PRIOR APPLICATION NUMBER: GB 0218518.9				
; PRIOR FILING DATE: 2002-09-08				
; PRIOR APPLICATION NUMBER: US 60/366,601				
; PRIOR FILING DATE: 2002-03-22				
; NUMBER OF SEQ ID NOS: 12				
; SOFTWARE: PatentIn version 3.2				
; SEQ ID NO 11				
; LENGTH: 3981				
; TYPE: DNA				
; ORGANISM: Mus musculus				
US-10-379-981-11				
Query Match 61.7%; Score 1487.6; DB 16; Length 3981;				
Best Local Similarity 79.7%; Pred. No. 0;				
Matches 1850; Conservative 0; Mismatches 419; Indels 51; Gaps 6;				
Qy	73	ATGACCATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTATCAGATCAGCCT	132	

Db	1	ATGACCATAGTTGACAAA	---ACTGAACCTTCAGACCCATCAACCTGTTCAGAACAGCCT	57
Qy	133	GGCAGCTCCGAGGAGTCTCACCTGGAGACATGGATGAGGTTCTGCCAGCTGGGGTGT	192	
Db	58	GGCAGTTGTGAGGGGGTCTCACCTGAACATGGAACAGAGGCTCTGCCAGCTGGGGCGT	117	
Qy	193	GTGTCTTCATTGAATGATGTCTAAATCACACACTTTCTTTAGGACCAAGTACTGGTGT	252	
Db	118	GTGTCTTCAATGAAGTGTCTCAAGTCACACACTTCCATTAGGGCCAGTGTGGTGT	177	
Qy	253	GTAGTTTATTCCAGTTCATCTGTACCTGATAAATCAAAACCATCACCAAAAGGATCAA	312	
Db	178	GTAGTTTATTCTAATCTGTCTGTACTGAAAAATCAAGGCCATCACCAAAAGGATCAA	237	
Qy	313	GCCCTAGTGTGATGCGATCGCTCTCCACAGAAAGTCTTTTCCCATCTGAGAAAGTGTGT	372	
Db	238	GTCTCTAGGTGTGCGATTTGCTCTCTCAAAAGGTCTCTGTCTCCATCTGAAAGATTTGT	297	
Qy	373	CTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGGCAACACCTGT	432	
Db	298	CTTAAGTGGCAACAAAGTCATCGAGTTGGCGCTGGGCTCCAGAAATTTGGGCAACACCTGT	357	
Qy	433	TTTGCAATGCAGCACTGCAGTGTTTAACTTACCTACACACACCTCTTTGCCAATTTACATGCTA	492	
Db	358	TTTGCAATGCAGCACTGCAGTGTCTGACTTACAGCCACCCCTCGCCCAATTTACATGCTA	417	
Qy	493	TCACATGAACACTCCCAAAACATGTCATGCAGAAAGCTTTTGTATGATGTGTACAATGAA	552	
Db	418	TCCCATGAACACTCCCAAGACATGCCACGAGAGGATTTTGTATGATGTGCAGATGCAG	477	
Qy	553	GCACATATTACCCAGGCACTCAGTAATCTCTGGGACGTTTATAACCAATGTTTCTCATC	612	
Db	478	ACACACATTA CCCAGGCACTTAGCAACCCCTGGGATGTTATCAAGCCGATGTTCTCATC	537	
Qy	613	AATGAGATGCGCGGTATAGCTAGGCACCTCCGTTTGGAAAACCAAGAAAGATGCCCATGAA	672	
Db	538	AATGAAATGCGCGGTATAGCTAGACACTTCGTTTGGAAACCAAGAAAGATGCCCATGAA	597	
Qy	673	TTCTCTCAATACACTGTTGATGCTATGAGAAAGCATGCTTGAATGGCAGCAATAAATTA	732	
Db	598	TTTCTTCAGTACACGGTGCATGCCATGCAGAAAGCATGTTTAAATGGCAGCAATAAATTA	657	
Qy	733	GACAGACACCCAGGCCACCACTCTGTTGCTCAGATATTGGGAGGATACCTAAGATCT	792	
Db	658	GACAGACACCCAGGCCACCACTCTGTTGCTCAGATATTGGGAGGATACCTAAGATCT	717	
Qy	793	AGAGTCAAAATGTTTAAATTTGCAAGGGCGTTTCAGATACCTTTTGATCCCATATCTTGATATA	852	
Db	718	CGAGTTAAATGTTTAAATTTGCAAGGGTGTTCAGATACCTTTTGATCCCATATCTTGACATA	777	
Qy	853	ACATTGGAGATAAAGGCTGCTCAGAGTGTCAAAGGGCAATGGAGAGTTTGTGAAGCGG	912	
Db	778	ACGTTGGAGATTAAGGCTGCACAGAGTGTTCACAGGGCTTAGAGCAGTTCGTGAAGCCA	837	
Qy	913	GAAACGCTGTGAGGAGAAACTCGTACAGTGCAGCAAGTGTAAAAGATGTTCCAGCT	972	
Db	838	GAAACACTGGATGGAGAAAACTCCACAAAGTGCAGAAAGTGCAGAAAAAATGTTCCAGCT	897	
Qy	973	TCAAGAGAGGTTTCACTATCCATAGATCTCTAATGTTCTTACACTTTCTCTGAAAACGTTTT	1032	
Db	898	TCAAGAGAGATTCACAAATCCATAGGTCCTCTAATGTTCTTACCATCTCACTGAAGCGCTTT	957	
Qy	1033	GCAAAATTTACCGGTGGAAAAATTCCTAAGGATGAAATACCCCTGAGTATCTTGATATT	1092	
Db	958	GCCAACTTCACCGGTGGAAAAATTCCTAAGGATGTGAAATATCCCTGAGTACCTTGATATC	1017	
Qy	1093	CGGCCATATATGCTCAACCCCAACGAGGACCAATGTCTAGCTTTGTATCGAGTGTG	1152	
Db	1018	CGGCCCATATATGCTCAGCCCAATGAGGACCAATTTATTTATGTTTGTATCTGTGCTG	1077	
Qy	1153	GTCCACACTGGTTTAAATTTGCCCATGCTGGCCATTAATCTCTGCTACATAAAAGTAGCAAT	1212	

Db 1078 GTGCACACTGGTTTAAATGTCTGTGGCCACTACTTTTGTCTACATCAAGGCTAGCAAT 1137
QY 1213 GGCTCTGGTATCAAAATGAATGACTCCATGTGTATCTACAGTATATATAGATCGGTACTC 1272
Db 1138 GGCTCTGGTATCAGATGAATGACTCCATGTGTCTCCACCAAGTATATCAGAGCAGTCTTT 1197
QY 1273 AGCCAAACAGCTATGTGCTCTTTTATATCAGGTGCCATGATGTGAAAAATGAGGTGAA 1332
Db 1198 AACCAAGAGCTTACGTGCTCTTTTATATCAGGTGCCATGATGTGAAAAATGAGGGGAG 1257
QY 1333 CTTACTCATCCCAACCCATAGCCCGGCGAGTCTCTCCCGCCCGGTCATCAGTCAAGCGG 1392
Db 1258 TCTGCTCATCTGCTCCATAGCCCGGCAATCTCTCCCGCCCGGAGTCAAGTCAAGCGG 1317
QY 1393 GTTGTCAACCAACAGGCTGCGCAGGCTTTATCGGACCAAGTCTTCCCTCTCAGATG 1452
Db 1318 GTAGTCAACCAACAGGCTGCGCAGGCTTTATCGGACCCAGCTGCTTCCCATGTG 1377
QY 1453 ATAAAGAACTCACTCACTTAAATGGGACTGGACCAATTGAAAGACACGCCAAGCAGTTCC 1512
Db 1378 ATGAAGAACAGCCACACACTTGAATGGCACCACGCCAGTGAAGACACACCAAGTAGTTCT 1437
QY 1513 ATGTGAGTCTTAAACGGGAATTCAGTGTCAACAGGGCTAGTCTGTGTTAATGTCTCAGCT 1572
Db 1438 GTGTCAAGCCCTTAACCGAAACACACAGCGTCAATAGGGCCAGTCTGTCTACTGCTTCGACT 1497
QY 1573 TCTGTCCAAACTGCTCAGTAAATAGTCTCTCAGTATCCAGAACATCCTAAGAACAA 1632
Db 1498 TCTGTGAGAACTGGTCTGTGTACAGACCCCTCAGTTAATCCAGATCACCCCAAGAAACAA 1557
QY 1633 AAAATTAACATCAGTATTCACAAAGTTCCTCTGTGCGCAGTGTCAAGTCTCAACCTAA- 1691
Db 1558 AAAATCACCATCAGTATTCACAAAGTTCCTCTGTGCGCAGGTCAGGACCACTGAAT 1617
QY 1692 -----CCTTCATAGTAAATCTTTGGAGAACCTTACCAAGCCGTTCCCTTCTTACCAATT 1746
Db 1618 AACAGGCTTCAATGGCCCTGTCTGTGGAGGCTCTCTAGTAAGCGCGCACCCCTCTCCACCATC 1677
QY 1747 ACCAA---TTCGTCACTACAGTCTACTCGAAGCATCTACGATGTCAGTTTCTAGTAA 1803
Db 1678 ACTAACCTTCTGCAATACAGTCTACTCGAAGCATCTACTCGAAGCATCTACTCGAAGCATCT 1723
QY 1804 GTAACAAAACCGATCCCGCAGTGAATCCTGTCTCCAGCCCGTGTGAATGCAAAATCC 1863
Db 1724 -----CGACTTCCCGCAGTGGCTGTCCCAAGCCCATGTGTGAAGCGCAGGCT 1773
QY 1864 AAGCTGAATCAGCGTGTGCTGTGCTTATGGCGCGAGTCTCTGTAGGACTCTGACGAG 1923
Db 1774 AAGTGGCGCCAGTGTGCTTGTCCCTATGGGCGCGAGTCTCTCAGAAAGTCTGATGAG 1833
QY 1924 GAGTCAAAAGGGCTGGCAAGGAGATGGATGTGTACGATTTGTAGCTCCCACTCTCCC 1983
Db 1834 GAGTCAAAAGGGCTGGCAAGGAGATGGATGTGTGTGATGATGATGATGATGATGATGATGAT 1893
QY 1984 GGCCAAAGATGC-----CGAAGATGAGGAGGCGCACTCCGACGAGCTTCAAGAACCC 2034
Db 1894 AGGCCAAGAGCTGCTGCAGATGAGCGTGTGTGAGGCTTCTCCCATGAGCTTCAAGAACCC 1953
QY 2035 ATGACCCCTAAACGGTCTTAATAGTGCAGACAGGACAGTGAACCCGAAAGAAACGGCCTA 2094
Db 1954 GTCTCTTAATAGTGTCTAATAGCGCA-----GACAGTGACTCACAAGAGAACAGCCTG 2007
QY 2095 GCGCCTGTAGTGTGCTGAGTGCCTGAGGCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2154
Db 2008 GCATTTGACAGTGCAGCTGCGAGTGCAGGCTGCGAGCGAGCTACACAGAAACCTCTTTTCC 2067
QY 2155 AAGGCAAAACGGTCTTCTGTGAAATGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2214
Db 2068 AAATCTTAATAGTGTCTTCTGTGAAAGTGTGAGCGCTGCTTCTTGTGAGTCTGCTGCTGAGAC 2127
QY 2215 AAATCTTACAGACCTTACAGCTTACCAAACTGAAAGGCTGCAAGGATGAATGAGT 2274
Db 2128 AGAATCTTGTAGACCTTCAAGCTTACCAACCGAGGCAAGGGTCCAGCGGTGAAGAGAGT 2187

QY 2275 GCACCTGGACAGAGGGGCGCTCCCGAGGACCGGACCGGAGCCTCAGCTGGCAGC 2334
Db 2188 TGGACTACGACAGAGGGGGAAGCTCTCCAAAGGACCCCTGTTTTCACAGCTGGAGCCCATCAGT 2247
QY 2335 CCGCGCCCGCAATCCCTGGAGGAGCAGATGCGGCGCGCAA 2374
Db 2248 GATGAGCCCAAGTCCCTTGGATACCGGAGGCTGTACCA 2287

RESULT 6

US-10-371-905A-33
; Sequence 33. Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 1593
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: human DUB8.7
US-10-371-905A-33

Query Match 18.4%; Score 443; DB 13; Length 1593;

Best Local Similarity 62.7%; Pred. No. 8e-132;

Matches 733; Conservative 0; Mismatches 415; Indels 21; Gaps 2;

QY 163 ATGGATGACAGGTTCTGCGAGCTGGGTGCTGTCTTTCAATTGAATGATGTCAAAATCAC 222
Db 1 ATGGAGGACGACTCACTCTACTTGGAGGTGAGTGGCAGTTCAACCACTTTTCAAAATC 60
QY 223 ACATTTCTTTAGGACCGAGTACCTGGTGTGTGTATTTATTCGAGTTTCATCTGACCTGAT 282
Db 61 ACATCTTCTGGCCAGATGCGAGCTTTTGTCTGAAATCCA---GGGACTTCTCTCCCTGAG 117
QY 283 AATCAAAAACCATCACCAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCTCTCCACAG 342
Db 118 AAGTACCACTCTCATCTGAGGCCGCTGTGCACTCTGTGATGATTTGGCTCTCTGTGGCA 177
QY 343 AAAGTTCTTTTCCCATCTGAGAAAGATTGTCTTAAAGTGGCAACAACTCATAGAGTTGGA 402
Db 178 AGACAGCTTGTCTCCAGGAGAGAGCTTCTCTGAGTAGAGGAGACCTGCTGCGTGGGG 237
QY 403 GCTGGGCTCCAGAAATTGGGCAATACCTGTTTGGCAATGACAGCTGCGAGTGTGTTAAAC 462
Db 238 GCTGGGCTCCAGAAATATGGAAATACCTGCTACGAGAACGCTTCCCTGCGAGTGCCTGACA 297
QY 463 TACACACCAAGCTTTTCCCAATTCATGCTCTATCATAGAACACTCCAAACATGTCATGCA 522
Db 298 TACACACCGCCCGCTTCCCAACTACATGCTGTCCGGGAGCACTCTCAAAATGTCAGCGT 357
QY 523 GAAGGCTTTTGTATGATGTGTACAATGCAAGACATATTTACCCAGGCACTCAGTAAATCTCT 582
Db 358 CCCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 417
QY 583 GGGAGCGTTTAAATCAATGTTTGTTCATCATGAGATGCGGGGTATAGTAGGCACTTC 642
Db 418 GGTCA-----TGTCAATCCAGCCCTCAGGCACTTGGCTGTGGGTTTC 459
QY 643 CGTTTTTGGAAACCAAGAGATGCCCATGAAATCTTCAATATACATCTGTGATGCTATGACAG 702
Db 460 CATAGAGGACGAGGAGAGATGCCCATGAAATTTCTCATGTTCTACTGTGATGCCATGAAA 519
QY 703 AAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCGCCACCACTCTTGT 762

Db 520 AAGGATGCTTCCGGCCACAGAGGTAGATCATCTAAGGACACACCCCTCATC 579
Qy 763 TGTGAGATATTTGGAGGATACCTAAGATCTAGAGTCAAAATGTTTAAATTTGAAGGGGCTT 822
Db 580 CACCAAAATATTTGGAGGCTGCTGGAGATCTCAAATCAAGTGTCTCCACTGCCACGGGATT 639
Qy 823 TCAGATCTTTTGTATCCATATCTTGTATATAATTTGGAGATAAAGGCTGCTCAGAGTGTG 882
Db 640 TCAGACACTTTTGAACCTTACCTGGACATCGCCCTGGATATCCAGGCGAGCTCAGAGTGTG 699
Qy 883 AACAGGCAATTTGGAGCAGTGTGTGAAGCGGGAACAGCTTGTATGGAGAAATCTCGTACAA 942
Db 700 AAGCMAGCTTTGGAAACAGTGTGTGAAGCCGGAAGAACTCAATGGAGAGATGCCCTATCAT 759
Qy 943 TGCAGCAAGTGTAAAGAGTGTCCAGCTTCAAGAGGTTCACATCATATAGATCCTCT 1002
Db 760 TGGGTCTTTGTCTCCAGAGGGCGCCAGCTCCAAGACGTTAACTTTACACACTTCTGCC 819
Qy 1003 AATGTTCTTACACTTCTCTGAAACGTTTGTGCAAAATTTTACGGTGGAAATTTGCTTAAG 1062
Db 820 AAGGTCTCTCATCTTGTCTTGAAGAGATTTCTCCGATGTACAGGCAACAAACTTGCCAAG 879
Qy 1063 GATGTGAATACCCCTGAGTATCTTCATATTTGGCCATATATGTCTCAACCCAAACGGAGAG 1122
Db 880 AATGTGCAATATCTCGAGTGCCTTGAATGCAATGCAAGCCATATCATGTCTCAGCAGNACACAGA 939
Qy 1123 CCAATGTCTAGCTTGTATGACGTGTGTCACACTGTTTAAATTTGAAGCCAGTGGC 1182
Db 940 CCTCTGTCTATGCTCTATGCTGTGTGTCACGCTGCGGTGGAGTGTCAOGATGGA 999
Qy 1183 CATTACTTCTGCTACATAAAGCTAGCAATGSCCTCTGCTATCAATGACTTCCATT 1242
Db 1000 CATTACTTCTCTTATGTCAAAGCTCAAGAAGCCAGTGGTATATAAATGGATGATGCCAAG 1059
Qy 1243 GTATCTACAGTATATAGATCGGTACTCAGCCAAACAGCCATGTGTCTCTTTTATATC 1302
Db 1060 GTCACTGCTGTAGCATCATCTTGTCTCTGATGATCAACAGGCCATGTCTCTTTTACATC 1119
Qy 1303 AGGTCCCATGATGTGAAATAAGAGGTGA 1331
Db 1120 CAGAGAGTGAATGGANAGACACAGTGA 1148

RESULT 7

US-10-371-905A-19
; Sequence 19, Application US/10371905A
; Publication No. US2003022496A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 19
; LENGTH: 1698
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: human DUB4.8
US-10-371-905A-19

Query Match 18.1%; Score 436; DB 13; Length 1698;
Best Local Similarity 62.1%; Pred. No. 1.6e-129;
Matches 732; Conservative 0; Mismatches 425; Indels 21; Gaps 2;
Qy 154 CTGGAGACATGGATCAGGTTCTGCGAGCTGGGGTGTGTCTTCAATTGAATGATGTG 213
Db 97 CCAGTCACATGGAGGACCACTCACTTACTTGGGAGGTGAGTGGCAGTTCAACCACTTT 156

Qy 214 TCBAATCACACACTTCTTTAGGACCAGTACCTGGTGTAGTATTTATTCGAGTTCACT 273
Db 157 TCAAAACTCATCTTCTCGGCCGATGCAGCTTTTGTGTAATCCA--GCGACTCTCT 213
Qy 274 GTACTGTATAAATCAAAACCAATCACCAAAAGGATCAAGCCCTAGAGTGTGATGCGCATCGCT 333
Db 214 CTCCTGAGAGTCAACCACTCTCATGTGAGACCCGTGTGACCTCTGTGATGATTTGGCT 273
Qy 334 CCTCCACAGAAAGTCTTTTCCCATCTCTGAGAGATTTCTCTTAAGTGGCAACAAACTCAT 393
Db 274 CCTGTGGCAACACAGCTTGTCTCCAGGGAGAAAGTTCCTCTCTAGTAGCAGGAGACCTGCT 333
Qy 394 AGATTGGAGCTGGCTCCAGAAATTTGGGCAATACCTGTTTGGCAATGCGACGACTGCGAG 453
Db 334 GCGGTGGGGCTGGCTCCAGAAATATGGAAATACCTCTAGTGAAGCTTCTCTTGGAG 393
Qy 454 TGTTTAACTCACACCACTCTTGGCCAAATTACATGTCTATCACATGAACACTTCCAAAACA 513
Db 394 TGCCTGACATACACACCGCCCTTGGCAACTACATGCTGTCCCGGAGCACTCTCAAAG 453
Qy 514 TGTATGCAAGAGGCTTTTGTATGATGTACATGCAAGCAATATTAACAGCACTC 573
Db 454 TGTATCTGTCAAGGGCTGCACTGCTCTGTACGATGCAAGCTCAATCATCACACGGGCCCTC 513
Qy 574 AGTAATCTCTGGGAGGTTTAAACCAATGTTTGTTCATCAATGAGATGGGGGTATAGCT 633
Db 514 CACAATCTCTGGCCAGCTCATCCAGCC-----CTCAGGCAATTTGGCT 555
Qy 634 AGGCACCTCCGTTTGGAAACCAAGAGATGCCATGAATTCCTTCAATFACACTGTGTAT 693
Db 556 GCTGCTTCCATAGAGGCAAGCAGGAAGATGCCATGAATTTCTCATGTTCACTGTGGAT 615
Qy 694 GCTATGCAAGCAAGCTTGTGAATGGCAGCAATAAATTAGACAGACACACCAGGCCAC 753
Db 616 GCCATGAAAAGGCAATGCTTCCCGGCACAAAGCAGGTAGATCATCTCTAAGGACACC 675
Qy 754 ACTCTGTGTGTGAGATATTTGGAGATACCTTAAGATCTAGAGTCAAAATGTTTAAATGTC 813
Db 676 ACCCTCATCCACCAATATTTGGAGGCTACTGAGATCTCAATCAAGTGTCTCCACTGC 735
Qy 814 AAGGGCTTTTCAGATACCTTTTGTATCCATATCTTGATATAACAAATTTGGAGATAAGGCTGT 873
Db 736 CACGGCAATTCAGACACTTTTGACCTTACCTGGACATCGCCCTGGATATCCAGGCACT 795
Qy 874 CAGAGTGTCAACAGGCAATTTGGAGCAGTTTGTGAAGCCGGAACAGCTTGTATCGAGAAAC 933
Db 796 CAGAGTGTCCAGCAAGCTTTGGAAACAGTTGGTGAAGCCGGAAGAACTCAATGGAGAGAT 855
Qy 934 TCGTACAAAGTGCAGCAAGTGTAAAGAGATGGTTCCAGCTTCAAAGAGGTTCACTATCCAT 993
Db 856 GCCTATCATTTGGTGTGTTGTTCTCCAGAGGGCGCGCCCTCCAGACGTTAATTTACAC 915
Qy 994 AGATCCTCTAATGTTCTTACACTTTCTCTGAAACGTTTTTGGAAAATTTTACCCTGGAAAA 1053
Db 916 ACCTCTGCCAAGGCTCTCATCTTGTATTGAAGAGATTTCTCCGATGTGACAGGCAACAG 975
Qy 1054 ATTGTAAAGATGTGAATACCTCAGTATCTTGTATTTGGGCCATATATGTCTCAACCC 1113
Db 976 ATTGCCAAGAAATGTGCAATATCTCTGAGTGCCTTGACATGTCAGCCATACATGCTCAGCAG 1035
Qy 1114 AAGGAGGACCAATTTGTCTACCTTGTATGAGTGTCTGTCCTCCACTGTTTAAATTCG 1173
Db 1036 AACACAGGACCTCTTGTCTATGTCTCTATGTGTGTGCTGAGTCCAGCTGGGTGGAGTGT 1095
Qy 1174 CATGCTGGCCATTACTTCTGTCTACATAAAGCTAGCAATGGCTCTGTGTATCAAAATGAT 1233
Db 1096 CACACGAGACATTAATCTTCTTATGTCAAGCTCAAGAGGCCAATGGTATATAAATGAT 1155
Qy 1234 GACTCCATTTGATATACCAAGTATATAGATCGGTACTCAGCCAAACAGCCATATGTGCTC 1293
Db 1156 GATCCGGAGTCAACCGCGCTAGCATCTTCTGTCTGAGTCAACAGGCTACCTGCTCTC 1215

QY 1294 TTTTATATCAGGTCCCATGATGTGAAATAATGAGGTGA 1331
|||||
Db 1216 TTTTATATCAGAGAGTGAATGGAAACACAGTGA 1253
|||||

RESULT 8
US-10-107-695B-1
; Sequence 1, Application US/10107695B
; Publication No. US2003002201A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 68999, Human Ubiquitin
; FILE REFERENCE: Carboxyl-Terminal Hydrolase Family Member and Uses Therefor
; CURRENT APPLICATION NUMBER: US/10/107,695B
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 60/279,184
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1763
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (171)....(1763)
US-10-107-695B-1

Query Match 18.0%; Score 434.6; DB 15; Length 1763;
Best Local Similarity 62.0%; Pred. No. 4.6e-129;
Matches 733; Conservative 0; Mismatches 429; Indels 21; Gaps 2;

QY 149 TCTCACTGAGACATGAGTGCAGGTTCTGCCAGCTGGGTGCTGTCTTCAATTGAATG 208
|||||
Db 157 TCGTTCCAGTCGACATGGAGGACACTCACTACTTTGGGAGGTGAGTGGCAGTTCAACC 216
|||||

QY 209 ATGTGTCAAAATCACACACTTCTTTAGGACCAGTACCTGTGTGTGTATTTATTCAGATT 268
|||||
Db 217 ACTTTTCAAAATCACACTTCTTCCGCCCGATGCAGCTTTTGTGAAATCCA---GCGGA 273
|||||

QY 269 CATCTGTACTGATATAAATCAAAACCAATCACCAAAAGGATCAAGCCCTAGGTGATGGCA 328
|||||
Db 274 CTCTCTCCCTGAGAGTACCACTCTCATGTGAGGCCGTGTGCGACCTCTGTGATGATT 333
|||||

QY 329 TCGCTCTCCACAGAAAGTCTTTTCCATCTCAGAGAGATTGTCTTAAGTGGCAACAA 388
|||||
Db 334 TGGCTCTCTGGCAAGACAGCTTGCTCCAGGAGAGACTTCTCTGAGTGAAGGAGAC 393
|||||

QY 389 CTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTTGGCAATGCGAGCAC 448
|||||
Db 394 CTGCTGGGTGGGGCTGGGCTCCAGAAATATGGAAATACCTGCTAGTGAACGCTTCT 453
|||||

QY 449 TGCAGTGTTTAACTTACACACACCTCTTGGCCAAATACATGCTATCACATGAAACATCCA 508
|||||
Db 454 TGCAGTGTCTGACATACACACCGCCCTTGGCCAAATACATGCTGTCCGGGAGCACTCTC 513
|||||

QY 509 AAACATGTCTATCAGAGGCTTTTGTATGATGTGTAATGCAATGCAATGATGATGCGGCTA 568
|||||
Db 514 AAACGTGTCTATGTCACAGGGCTGTATGCTGTATGCAAGCTCATCATCACCGGG 573
|||||

QY 569 CACTCATGTAATCTGGGGAGCTTTATTAACCAATGTTTGTTCATCAATGATGATGCGGCTA 628
|||||
Db 574 CCCTCCACAAATCTGGCCAC-----GTCACTCCAGCCCTCACAGGCAT 615
|||||

QY 629 TAGCTAGGACCTCCGTTTGGAAACCAAGAGATGCCCAATGAATTCCTTCAATACATG 688
|||||
Db 616 TGGCTCTGCTGCTTCCATAGAGGCAAGCAGAGATGCCCAATGAATTTCTCATGTTCACTG 675
|||||

QY 689 TTGATCTATGCAAGAGAGTGTGTAATGGCAGCAATAAATTTAGACAGACACACCCAGG 748
|||||
Db 676 TGGATGCCATGAATAAGGCAATGCTTCCCGGGCAAGCAGGTGATGATCATCACTTAAGG 735
|||||

QY 749 CCACCACCTCTGTTTGTTCAGATATTTGGAGATACCTAAGATCTAGAGTCAAAATGTTTAA 808
|||||
Db 736 ACACCACCCTCATCCACCAATATTTGGAGGCTACTGGAGATCTCAAAATCAAGTGTCTCC 795
|||||

QY 809 ATTGCAAGGGGCTTTTCAGATACCTTTTTCATCATATCTTGATATAAATTTGGAGATAAAGG 868
|||||
Db 796 ACTGCCACGGCATTTTCAGACACATTTTGACCTTACCTTGGACATCGCCCTGGATATCCAGG 855
|||||

QY 869 CTGCTCAGAGTGTCAACAAGGCATTTGGAGCAGTTTGTGAAGCCGGAACACAGCTTGAATGAG 928
|||||
Db 856 CAGCTCAGAGTGTCCAGCAAGCTTTTGAACAGTTTGTGAAGCCCGCAAGAACTCAATGGAG 915
|||||

QY 929 AAAACTCGTACAAGTGCAGCAAGTGTAAAGATGTTTCCAGCTTCAAGAGGTTTCACTA 988
|||||
Db 916 AGAATGCTATCATTTGGGTGTTTCTCCAGAGGGCGCCGCTCCCAAGACGTTAACTT 975
|||||

QY 989 TCCATAGATCCTCTTAATGTTCTTACACTTTTCTCTGAAACGTTTTTGCAAAATTTTACCGGTG 1048
|||||
Db 976 TACACACCTCTGCCAAGTCTCATCTTCTTAATGAAGAGATTCTCCGATGTACAGGCA 1035
|||||

QY 1049 GAAAAATTGCTAAGGATGTGAATACCTGAGTATCTTGATATTCGGCCATATATGTCTC 1108
|||||
Db 1036 ACAAGATTGCCAAGATGTGAATATCTGAGTGCCTTGACATGCAGCCCATACATGTCTC 1095
|||||

QY 1109 AACCAAGGAGAGGCAATTTGTCTAGCTCTGTATGTCAGTGTGCTCCACACTGGTTTAA 1168
|||||
Db 1096 AGCAGAACACAGGACCTTGTCTATGCTCTATGCTGTGCTGCTCCAGCTGGGTGGA 1155
|||||

QY 1169 ATTGCCATGTGGCCATTTACTTCTGTCTACATAAAAGCTAGCAATGGCCTCTGGTATCAAA 1228
|||||
Db 1156 GTTGTCAACAGCAGCATTTACTTCTTATGTCAAAGCTCAAGAAGGCCAGTGGTATAAAA 1215
|||||

QY 1229 TGAATGACTCCATTTGTATCTACAGTGATATTAGATCGGTACTCAGCCCAACAGCCTATG 1288
|||||
Db 1216 TGGATGATCGGAGTCAACGCTCTAGCATCACTTCTGCTCTGAGTCAACAGGCTTACG 1275
|||||

QY 1289 TGCTCTTTTATATCAGTCCCATGATGTGAAATTTGAGGTGA 1331
|||||
Db 1276 TCCTCTTTTATCCAGAGAGTGAATGGGAAAGACACAGTGA 1318
|||||

RESULT 9
US-10-371-905A-17
; Sequence 17, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 1593
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: human DUB4.7
US-10-371-905A-17

Query Match 18.0%; Score 433.4; DB 13; Length 1593;
Best Local Similarity 62.2%; Pred. No. 1e-128;
Matches 727; Conservative 0; Mismatches 421; Indels 21; Gaps 2;

QY 163 ATGGATGAGGTTCTCCAGCTGGGGTGTGTCTTTCATTTGAATGATGTGTCAAATCAC 222
|||||
Db 1 ATGGAGGACGACTCACTCTACTTGGGTGGTGGAGTGCAGTTCACCCACTTTTCARAACTC 60
|||||

QY 223 ACATTTCTTTTAGGAACAGTACCTGGTGTGTAGTTTATTCGAGTTTCACTGTACCTGAT 282
|||||

Db 61 ACATCTTCTCGCCCGGATGCGAGCTTTTGTGTAATCCA---CGGACTTCTCTCCTCGAG 117
QY 283 AAATCAAAACCATACCAAAAGGATCAAGCCCTTAGGTGATGGCATCGCTCTCCACAG 342
Db 118 AAGTCACCACCTCTCATGTGAGACCGGTGCGACCTCTGTGATGATTTGGCTCTCTGGCA 177
QY 343 AAAGTTCTTTTCCCATCTGAGAGATTTGTCTTAAGTGGCAACAACCTCATAGAGTTGA 402
Db 178 AGACAGCTGTCTCCAGGGGAAGCTTCTCTGAGTAGCAGGACCTGCTCGGTGGG 237
QY 403 GCTGGCTCCAGAAATTTGGCAATACCTGTTTGTGCAATGAGCACTGCGAGTTTAAAC 462
Db 238 GCTGGCTCCAGAAATTTGGCAATACCTGTTTGTGCAATGAGCACTGCGAGTTTAAAC 297
QY 463 TACACACACCTCTTGGCAATTTACATGTTATCACATGAACATCCCAAAACATGTATGCA 522
Db 298 TACACACCGCCCTTGCACACTACATGCTGTCCCGGAGCACTCTCAAAAGTGTCTATGT 357
QY 523 GAAGCTTTTGTATGATGTGTAATGCAATGCAAGCACATATTTACCCAGGCACTCAGTAATCT 582
Db 358 CACAAGGCTGATGCTCTGTACGATGCAAGCTCAGTCAATCAGACCGGCCCTCCACATCT 417
QY 583 GGGAGCGTTATTAACCAATGTTGTCTATCAATGAGATGCGCGTATAGCTAGGCACCTC 642
Db 418 GGCCAGCTCATCCAGC-----CTCACAGGCAATGGCTGCTGCTGC 459
QY 643 CGTTTTGGAAACCAAGATGCCATGAATTCCTTCAATA CACTGTTGATGCTATGCG 702
Db 460 CATAGAGGCAAGCAAGATGCCATGAATTTCTCATGTTCCAGTGTGATGCCATGAA 519
QY 703 AAAGCATGCTTGAATGGCAGCAATAAATAGACAGACACACCGGCCACCACTCTTGT 762
Db 520 AAGGTCAGCTTCCCGGGCACAAGCAGGTAGATCATCTCAAGACACCACTCTATC 579
QY 763 TGTCAATATTTGGAGGATCACCTAAGATCTAGAGTCAAAATGTTTAAATGCAAGGCGTT 822
Db 580 CACCAAAATTTGGAGGCTACTGGAGATCAATCAACTGTCTCCACTGCCACGGCAAT 539
QY 823 TCAGATATTTTGTATCCATCTTGAATTAACATTTGGAGATAAAGCTGCTCAGAGTGT 882
Db 640 TCAGACATTTTGAACCTTACCTGACATCGCCCTGGATATCCAGGCACTCAGAGTGT 699
QY 883 AACAGGATTTGGACAGCTTTGTGAAGCGGACACAGCTGTAGTGGAGAACTCTGACAG 942
Db 700 CAGCAAGCTTTGGACAGCTTTGTGAAGCGGACCAAGAACTCAATGGAGAGATGCCATCAT 759
QY 943 TGCAGCAAGTGTAAAGATGTTTCAGCTTCAAGAGGTTCACTATCCATAGATCCTCT 1002
Db 760 TGTGTTGTTGTTCTCCAGAGGCGCGGCTCCAGACGTTAACTTTACACACCTCTGCC 819
QY 1003 AATGTTCTTACACTTCTCTGAAAAGTTTGGCAATTTTACCGGTGGAAAAATGCTAAG 1062
Db 820 AAGGTCCTCATCTTGTATTTGAAGAGATTTCTCGATGTACAGGCAACAAGATTTGCAAG 879
QY 1063 GATGTGAATACCTTGAGTATCTGATATTCGGCCATATCTCTCAACCCCAACGGAGAG 1122
Db 880 AATGTGCAATATCTTGAAGTCTTGCATGTGAGGCCATATATGTTCTCAGCAGAACACAG 939
QY 1123 CCAATTTGTCTAGCTCTTGTATGAGTGTGTTCCACACTGGTTTTAAATGGCAATGCTGGC 1182
Db 940 CTTCTTGTATGTTCTCTATGTTGTTGTTGTTCCAGCTGGGTGGAGTTGTTCACACGGA 999
QY 1183 CATTAATCTTGTCTACATAAAGCTAGCAATGGCTCTGTGTATCAATGAATGACTCCATT 1242
Db 1000 CATTAATCTTCTTATGTCAAGCTCAAGAGGCGCAGTGGTATATAAATGATGATGCCGAG 1059
QY 1243 GTATCTACAGTATATAGATGGTACTCAGCCACAGACCTATGTGCTCTTTTATATC 1302
Db 1060 GTACCGCGCTAGCATCACTTGTGCTGAGTCAACAGGCGCTACGTCCTCTTTTACATC 1119
QY 1303 AGGTCCCATGATGTGAAAAATGGAGTGA 1331
Db 1120 CAGAGAGTGAATGGGAAAGACACAGTGA 1148

RESULT 10
US-10-371-905A-11
; Sequence 11, Application US/10371905A
; Publication No. US2003022496A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 1593
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: human DUB4.3
US-10-371-905A-11

Query Match 17.9%; Score 431.8; DB 13; Length 1593;
Best Local Similarity 62.1%; Pred. No. 3.5e-128;
Matches 726; Conservative 0; Mismatches 422; Indels 21; Gaps 2;
QY 163 ATGGATGCGAGGTTCTGCGCAGCTGGGGTGTGTCTTCAATGAATGATGTCAAAATCAC 222
Db 1 ATGGAGGAGCACTCACTTACTTGGAGGTGAGTGGCAGTTCACCACTTTTCAAAACTC 60
QY 223 ACACCTTTCTTTAGGACCGAGTACCTGGTGTGTGATTTTATTCAGTTCATCTGTACTGTAT 282
Db 61 ACATCTTCTCGGCCCGGATGCGAGCTTTTGTGTAATCCA---CGGACTTCTCTCCTCGAG 117
QY 283 AAATCAAAACCATACCAACAAAGGATCAAGCCCTTAGGTGATGGCATCGCTCTCCACAG 342
Db 118 AAGTCACCACCTCTCATGTGAGACCGGTGTCGACCTCTGTGATGATTTGGCTCTCTGGCA 177
QY 343 AAAGTTCTTTTCCCATCTGAGAGATTTGTCTTAAGTGGCAACAACCTCATAGAGTTGA 402
Db 178 AGACAGCTGTCTCCAGGGGAAGCTTCTCTGATGAGCAGGAGACCTGCTCGGTGGG 237
QY 403 GCTGGCTCCAGAAATTTGGCAATACCTGTTTGTGCAATGAGCACTGCGAGTTTAAAC 462
Db 238 GCTGGCTCCAGAAATTTGGCAATACCTGTTTGTGCAATGAGTGGCGGTATAGTAGGCACCTC 297
QY 463 TACACACACCTTGGCCAAATTTACATGCTATCACATGACACTCCCAAAACATGTATGCA 522
Db 298 TACACACCGCCCTTGGCCAACTACATGCTGTCCCGGAGCACTCTCAACAGTGTATGCT 357
QY 523 GAAGCTTTTGTATGATGTGTAATGCAATGCAAGCACATATTTACCCAGGCACTCAGTAATCT 582
Db 358 CACAAGGCTGATGCTCTGTACGATGCAAGCTCAGTCAATCAGACCGGCCCTCCACATCT 417
QY 583 GGGAGCGTTATTAACCAATGTTGTCTATCAATGAGATGCGCGTATAGTAGGCACCTC 642
Db 418 GGCCAGCTCATCCAGC-----CTCACAGGCAATGGCTGCTGCTGC 459
QY 643 CGTTTTGGAAACCAAGATGCCATGAATTCCTTCAATA CACTGTTGATGCTATGCG 702
Db 460 CATAGAGGCAAGCAAGATGCCATGAATTTCTCATGTTCCAGTGTGATGCCATGAA 519
QY 703 AAAGCATGCTTGAATGGCAGCAATAAATAGACAGACACACCGGCCACCACTCTTGT 762
Db 520 AAGGTCAGCTTCCCGGGCACAAGCAGGTAGATCATCTCTAAGACACCACTCTCATC 579
QY 763 TGTCAATATTTGGAGGATCACCTAAGATCTAGAGTCAAAATGTTTAAATGCAAGGCGTT 822
Db 580 CACCAAAATTTGGAGGCTACTGGAGATCAATCAACTGTCTCCACTGCCACGGCAAT 539
QY 823 TCAGATATTTTGTATCCATCTTGAATTAACATTTGGAGATAAAGCTGCTCAGAGTGT 882
Db 640 TCAGACATTTTGAACCTTACCTGACATCGCCCTGGATATCCAGGCACTCAGAGTGT 699
QY 883 AACAGGATTTGGACAGCTTTGTGAAGCGGACACAGCTGTAGTGGAGAACTCTGACAG 942
Db 700 CAGCAAGCTTTGGACAGCTTTGTGAAGCGGACCAAGAACTCAATGGAGAGATGCCATCAT 759
QY 943 TGCAGCAAGTGTAAAGATGTTTCAGCTTCAAGAGGTTCACTATCCATAGATCCTCT 1002
Db 760 TGTGTTGTTGTTCTCCAGAGGCGCGGCTCCAGACGTTAACTTTACACACCTCTGCC 819
QY 1003 AATGTTCTTACACTTCTCTGAAAAGTTTGGCAATTTTACCGGTGGAAAAATGCTAAG 1062
Db 820 AAGGTCCTCATCTTGTATTTGAAGAGATTTCTCGATGTACAGGCAACAAGATTTGCAAG 879
QY 1063 GATGTGAATACCTTGAGTATCTGATATTCGGCCATATCTCTCAACCCCAACGGAGAG 1122
Db 880 AATGTGCAATATCTTGAAGTCTTGCATGTGAGGCCATATATGTTCTCAGCAGAACACAG 939
QY 1123 CCAATTTGTCTAGCTCTTGTATGAGTGTGTTCCACACTGGTTTTAAATGGCAATGCTGGC 1182
Db 940 CTTCTTGTATGTTCTCTATGTTGTTGTTGTTCCAGCTGGGTGGAGTTGTTCACACGGA 999
QY 1183 CATTAATCTTGTCTACATAAAGCTAGCAATGGCTCTGTGTATCAATGAATGACTCCATT 1242
Db 1000 CATTAATCTTCTTATGTCAAGCTCAAGAGGCGCAGTGGTATATAAATGATGATGCCGAG 1059
QY 1243 GTATCTACAGTATATAGATGGTACTCAGCCACAGACCTATGTGCTCTTTTATATC 1302
Db 1060 GTACCGCGCTAGCATCACTTGTGCTGAGTCAACAGGCGCTACGTCCTCTTTTACATC 1119
QY 1303 AGGTCCCATGATGTGAAAAATGGAGTGA 1331
Db 1120 CAGAGAGTGAATGGGAAAGACACAGTGA 1148

Db 640 TCAGACACTTTTACCTTACCTGACATCGCCCTGGATATCCAGGAGCTCAGAGTGC 699
QY 883 AACAGGCAATGAGAGCTTTGTAAGCCGGAACAGCTTGAATGAGAGAAACTCGTACAAG 942
Db 700 CAGCAAGCTTTGGAACAGCTTTGGTGAAGCCGGAAGAACTCAATGGAGAGAAATGCCATCAT 759
QY 943 TGCAGCAAGGTGAAGAGATGTTCCAGCTTCAAGAGGTTCACTATCCATAGATCTCT 1002
Db 760 TGTGGTGTGTTCTCCAGAGGGCCCGCCCTCCAAAGACGTTAACTTTACACACCTCTGCC 819
QY 1003 AATGTTCTTACACTTCTCTGAAACGTTTGGCAAAATTTTACCGGTGGAATAATGCTTAAG 1062
Db 820 AAGTCTCATCTTGTATTTGAAGAGATTTCTCCGATGTGACAGCAACAGATTTGCCAAG 879
QY 1063 GATGTGAATATACCTGAGTATCTTGATATTTGGCCCATATATGTCTCAACCAACGAGAG 1122
Db 880 AATGTGCAATATCTGAGTGCCTTGACATGAGCCATACATGTCTCAGCAGAACACAGGA 939
QY 1123 CCAATTGTCTACGTTCTGTATGAGTCTGCTGCTCCACACTGTTTAAATGCCATGCTGCC 1182
Db 940 CCTCTGTCTATGTCTTATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 999
QY 1183 CATTTACTTCTGTACATAAAGCTAGCAATGGCTCTGTGATCAAAATGAATGACTCCATT 1242
Db 1000 CATTTACTTCTGTATGTCAAGCTCAAGAAAGCCAAATGGTATAAATGGATGATGCCGAG 1059
QY 1243 GTATCTACAGTATATAGATGGTACTCAGCCCAACAGCTATGTGCTCTTTTATATC 1302
Db 1060 GTACCGCCCTGACATCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1119
QY 1303 AGTCCCATGATGTGAATAATGGAGTGA 1331
Db 1120 CAGAAGAGTGAATGGGAAGACACAGTGA 1148

RESULT 11

US-10-371-905A-3
; Sequence 3, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Hong
; APPLICANT: Hahn, Chang
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0195
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 3051
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: human DUB4.1a
US-10-371-905A-3

Query Match 17.9%; Score 431.2; DB 13; Length 3051;
Best Local Similarity 61.9%; Pred. No. 9e-128;
Matches 729; Conservative 0; Mismatches 428; Indels 21; Gaps 2;
QY 154 CTGGAGACATGATGACAGTCTTCCAGCTGGGGTCTGTGCTTCAATGAATGATG 213
Db 1450 CAGTGCATATGAGAGCAGTCACTACTTCTGAGAGTGAAGTGGCAGTTCAACCACTTT 1509
QY 214 TCAAACTCACACATTTCTTTAGGACCAAGTACCTGGTCTGTAGTTTATTCGAGTTCACT 273
Db 1510 TCAAACTCACATCTTCTCGCCAGATGACGCTTTTCTGCAAAATCCA---GCGTACTTCT 1566
QY 274 GTACCTGATTAATCAAAACCATCAACAAAGGATCAAGCCCTAGGTGATGCGATGCT 333
Db 1567 CTCCTTGAGAAAGTCAACCACTCTCATGTGTGAGACCGGTGTGACCTCTGTGATGATTTGGCT 1626

QY 334 CCTCCACAGAAAGTTCTTTTCCATCTGAGAAAGATTTGTCTTAAGTGGCACAACTCAT 393
Db 1627 CTTGTGGCAAGACAGCTTTGTCTCCAGGGAGAAAGCTCTCTGTAGTACAGAGAGACCTGTCT 1686
QY 394 AGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACACCTGTTTGGCAATGAGCACTCGAG 453
Db 1687 GCGTGGGGCTGGCTCCAGAAATTTGGGAAATACCTGCTACGTGAAGCTTCCCTGGAG 1746
QY 454 TGTTTAACTTACACACCACTCTTGGCAATTAATGCTATCACATGAACACTTCCAAAACA 513
Db 1747 TGCTGACATACAAAACGCGCACTTGCCTCATGCTGTTCCGGGAGCACTCTCAAAACG 1806
QY 514 TGTTCATGCAAGAGCTTTTGTATGATGTGTACATGCAAGCAATATATCCAGGCACTC 573
Db 1807 TGTTCATGCTCACAGGCTGCTGCTGTACTATGCAAGCTCACATCAAGAGGCGCTC 1866
QY 574 AGTAAATCCTGGGACGTTTATTAACCAATGTTTGTCTATCAATGAGATGCGCGCTATAGCT 633
Db 1867 CACATTCCTGGCCA-----TGTCAATCCAGCCCTCACAGGCAATGGCT 1908
QY 634 AGGACCTCGTGTGGAAACCAAGAGATGCCCATGAATTCCTTCAATACACTGTTGAT 693
Db 1909 GCTGGCTTCCATAGAGCAAGCAGAGAGATGCCCATGAATTTCTCATGTTCACTGTGGAT 1968
QY 694 GCTATGCAAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACCCAGGCCACC 753
Db 1969 GCCATGAGAAAGCATGCTTCCCGGCAACAGAGGTAGATCGTCACTCTAAGGACACC 2028
QY 754 ACTCTTGTGTGATGATATTTGGAGGATACCTAAGATCTAGAGTCAAAATGTTTAAATGTC 813
Db 2029 ACCCTCATCCACCAATATTTGGAGGCTACTGGAGATCTCAATCAAGTGTCTCCACTGC 2088
QY 814 AAGGGCTTTCAGATCTTTTGTATTCATATCTTGTATATAACATTTGGAGATAAAGGCTGCT 873
Db 2089 CACGGCAATTTACAGACATTTTGACCTTACCTGGACATCGCCCTGGATATCCAGGCACT 2148
QY 874 CAGAGTGTCAACAGGATTTGGAGGATTTGTAAGCCGGAACAGCTTGTATGAGAGAAAC 933
Db 2149 CAGAGTGTCCAGCAGCTTTGGAAACAGTGTGTGAGCCCGAGAACTCAATGGAGAGAT 2208
QY 934 TCGTACAAGTGCAGCAAGTGTAAAAAGATGTTTCCAGCTTCAAGAGGTTTCACTATCCAT 993
Db 2209 GCCTATCATTTGTGTGTTGTTCTCCAGAGGGCGCGCCCTCCAGAGGTTTAACTTTACAC 2268
QY 994 AGATCCTCTAATGTTCTTACACTTCTCTGAAAGCTTTTGCATAATTTTACGGTGGAAAA 1053
Db 2269 AACTCTGCCAAGGTCCTCATCTTGTATTGAAGAGATTTCCCGATGTCAAGGCAACAAA 2328
QY 1054 ATTGCTAAGGATGTGAAATACCTGAGTATCTTGTATATTCGGCCATATATGTCTCAACCC 1113
Db 2329 ATTGCCAAGATGTGCAATATCTGAGTGCCTTGAATGAGCCATACATGTCTCAGCAG 2388
QY 1114 AACGGAGAGCAATTTGTCTACGCTTTGTATGAGTGTGTTCCACACTGGTGTAAATGTC 1173
Db 2389 AACACAGGACCTCTCTGCTATGCTCTATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2448
QY 1174 CATGCTGGCCATTTACTTCTGCTACATAAAGCTAGCAATGGCTCTGCTATCAATGAAT 1233
Db 2449 CACAACGGACATTTACTCTCTTATGTCAAAGCTCAAGAGGCGCAGTGGTATATAAATGGAT 2508
QY 1234 GACTCCTATTGTATCTACAGTATATTAGATCGTACTCAGCCCAACAAAGCTATGTGCTC 1293
Db 2509 GATGCCAGGTACCGCTCTAGCANTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2568
QY 1294 TTTTATATCAGGTCCCATGATGTGAAAAATGGAGGTGA 1331
Db 2569 TTTTATATCCAGAGAGTGAATGGGAAAGACACAGTGA 2606

RESULT 12

US-10-107-695B-3
; Sequence 3, Application US/10107695B


```
; Publication No. US20030022201A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 68999, Human Ubiquitin
; FILE OF INVENTION: Carboxyl-Terminal Hydrolase Family Member and Uses Therefor
; FILE REFERENCE: MP101-076P1RCP1M
; CURRENT APPLICATION NUMBER: US/10/107,695B
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 60/279,184
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1590
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1590)
US-10-107-695B-3

Query Match      17.8%; Score 430.2; DB 15; Length 1590;
Best Local Similarity 62.0%; Pred. No. 1.1e-127;
Matches 725; Conservative 0; Mismatches 423; Indels 21; Gaps 2;

QY 163 ATGGATGAGGTTCTGCGAGCTGGGTGCTGTGCTTCAATGAAATGATGTGTCAAATCAC 222
DB 1 ATGGAGGAGCACTCACTCTACTTGGGAGGTGAGTGGGAGTCAACCACTTTTCAAAATC 60
QY 223 ACACCTTTCTTAGGACCACTACCTGCTGTGCTGTAGTTTATTCGAGTTTCACTGTACCTGAT 282
DB 61 ACATCTTCTCGGCCGATGTCAGCTTTTGTCTGAATCCA---GCGACTTCTCTCCCTGAG 117
QY 283 AAATCAAAACCATCACCAAAAGATCAAGCCCTAGTGTGATGGCATCGCTCTCCACAG 342
DB 118 AAGTCAACCACTCTCATGTGAGACCCGCTGTCACCTCTGTGATGATTTGGCTCCTGTGCA 177
QY 343 AAAGTTCTTTTCCCATCTGAGAAATGTTGCTTAAATGCTATCTGATGCTGATGCTGATGCTG 402
DB 178 AGACAGCTTGTCCAGGAGAAAGCTTCTCTGAGTACGAGAGACCTGCTGCGGTGGG 237
QY 403 GCTGGCTCCAGAAATTTGGGCAATACCTGTTTGGCAATGAGCACTGAGTGTGTTAACC 462
DB 238 GCTGGCTCCAGAAATGAGAAATACCTGCTACGTGAAGCTTCTTGGAGTGCCTGACA 297
QY 463 TACACACCACTCTTGGCAATTAAGTATACATGAACACTCCAAACATGTGATGCA 522
DB 298 TACACACCGCCCTTGGCAACTACATGCTGTCCCGGAGCACTCTCAAAACGTGTGATCGT 357
QY 523 GAAGGCTTTTGTATGATGTGATGATGCAAGCAATATACCAGGCACTCAGTAACTCT 582
DB 358 CACAGGCTCGATGCTCTGATGATGCAAGCTTCACTACACAGGCGCTCCACAACTCT 417
QY 583 GGGAGCTTTATTAACCAATGTTTGTATCAATGATGCGGCGTATAGCTAGGCACTC 642
DB 418 GGCAAGCTATCCAGCC-----CTCACAGGCAATGGCTGCTGGCTTC 459
QY 643 CGTTTGGAAACCAAGAGATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGAG 702
DB 460 CATAGAGGCAAGCAGGAAGATGCCCATGAATTTCTCATGTTCACTGTGGATGCCATGAA 519
QY 703 AAAGNTGCTTGAATGGAGCAATTAATTAAGACACACACCCAGGCACTCTTGT 762
DB 520 AAGGATGCTTCCCGGCAAGCAGGATGATATCATCTCTTAAGGACACCACTCTATC 579
QY 763 TGTCAATATTTGGAGGATACCTAAGATCTAGAGTCAATGTTTAAATTTGAAGGGCGTT 822
DB 580 CACCAATATTTGGAGGCTACTGGAGATCTCAATCAAGTGTCTTCACTGCAAGGCAAT 639
QY 823 TCAGATATTTTGAATCCATATCTGATATAACATTTGGAGATAAAGGCTGCTCAGAGTGT 882
DB 640 TCAGACACTTTTGAACCTTACTGAGACATCGCCTGGATATCCAGGCACTCAGAGTGT 699
```

```
QY 883 AACAGGCAATGGAGCAGTTTGTGAAGCCGGAACAGCTTGTGAGAGAAACCTCGTACAG 942
DB 700 CAGCAAGCTTTGGAAACAGTTGGTGAAGCCCGAAGAACTCAATGGAGAGATGCCTATCAT 759
QY 943 TGCAGCAAGTGTAAAGAGATGGTTCAGCTTCAAGAGAGTTTCACTATCCATAGATCCTCT 1002
DB 760 TGTGGTGTTCCTCCAGAGGGCGCGGCTCCAGAGAGTTAACTTTACACACCTCTGCG 819
QY 1003 AATGTTCTTACACTTTCTCTGAAAGCTTTTGAATTTTACCGTGGGAAATTTGCTTAA 1062
DB 820 AAGTCTCTCATCCTTGTATTGAAGAGATTCTCCGATGTACAGGCAACAAGATTGCCAAG 879
QY 1063 GATGTGAATACCTTGATATCTTGATATTCGCCATATATGTCACCCCAACGAGAG 1122
DB 880 AATGTGCAATATCTTGATGCTTGAATGAGCAGCATCATGCTTCACAGAAACACAGA 939
QY 1123 CCAATTTGTCTAGCTTCTGTATGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1182
DB 940 CCTCTGTCTATGCTCTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 999
QY 1183 CATTACTTCTGCTACATAAAAGCTAGCAATGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1242
DB 1000 CATTACTTCTCTTATGTCAAAGCTCAGAGAGCCAGTGGTATATAAATGGATGCGGAG 1059
QY 1243 GTATCTACCACTGATATAGATCGGTACTCAGCCCAACAGCTATGCTCTTTTATATC 1302
DB 1060 GTCACCGCTCTAGCATCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1119
QY 1303 AGGTCCCATGATGTGAATAATGGAGTGA 1331
DB 1120 CAGAAGAGTGAATGGAAAGACACAGTGA 1148
```

RESULT 13

```
US-10-371-905A-15
; Sequence 15, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 1593
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: human DUB4.6
US-10-371-905A-15
```

```
Query Match      17.8%; Score 430.2; DB 13; Length 1593;
Best Local Similarity 62.0%; Pred. No. 1.1e-127;
Matches 725; Conservative 0; Mismatches 423; Indels 21; Gaps 2;

QY 163 ATGGATGAGGTTCTGCGAGCTGGGTGCTGTGCTTCAATGAAATGATGTGTCAAATCAC 222
DB 1 ATGGAGGAGCACTCACTCTACTTGGGAGGTGAGTGGGAGTCAACCACTTTTCAAAATC 60
QY 223 ACACCTTTCTTAGGACCACTACCTGCTGTGCTGTAGTTTATTCGAGTTTCACTGTACCTGAT 282
DB 61 ACATCTTCTCGGCCGATGTCAGCTTTTGTCTGAATCCA---GCGACTTCTCTCCCTGAG 117
QY 283 AAATCAAAACCATCACCAAAAGATCAAGCCCTAGTGTGATGGCATCGCTCTCCACAG 342
DB 118 AAGTCAACCACTCTCATGTGAGACCCGCTGTCACCTCTGTGATGATTTGGCTCCTGTGCA 177
QY 343 AAAGTTCTTTTCCCATCTGAGAAATGTTGCTTAAATGCTAAGTGGCAACAACTCATAGAGTGT 402
```

Db 178 AGACAGCTTGCTCCAGGAGAGAGCTTCTCTCAGTAGCAGGAGACCTGCTCGGTGGG 237
QY 403 GCTGGGCTCAGAAATTTGGGCAATACCTGTTTGGCCAAATGCAGACACTGCAAGTGTAAAC 462
Db 238 GCTGGGCTCAGAAATATGGAAATACCTGCTAGCTGAACGCTTCTTGCAGTGCCTGACA 237
QY 463 TACACACCACTCTTGCCTAATCATGCTATCATCATGAACACTCCAAACATGCTATGCA 522
Db 298 TACACACCGCCCTTGGCAACTACATGCTCTCCGGGAGACACTCTCAACGCTGCTATCGT 357
QY 523 GAAGGCTTTTGATGATGTGTACAAATGCAAGCACAATATTAACCCAGGCACTCAATATCCT 582
Db 358 CACAAGGCTGTATGCTCTGCTAGCATGCAAGCTCACATCACAGGCGCTCCCAATCCT 417
QY 583 GGGGAGCTTATTAACCAATGTTTGTTCATCAATGAGATGGGCTATAGTAGGACCTC 642
Db 418 GGCACGCTCATCAGCC-----CTCACAGGCAATGGCTGCTGCTTC 459
QY 643 CGTTTTTGGAAACCAAGAGATGCCATGAATTCCTTCAATACACTGTGTGATGCTATGCGAG 702
Db 460 CATAGAGGCAAGCAGAGAGATGCCATGAATTTCTCATGTTCACTGTGATGCCATGAA 519
QY 703 AAGGATGCTGTGAATGGCAGCAATAAATATAGACAGACACACCCAGGCACTCTTTGTT 762
Db 520 AAGGCTATGCTTCCCGGCAACAGCAGGTGATCATCACTCTAAGGACACACCCCTCATC 579
QY 763 TGTTCAGATTTTGGAGATACCTAAGATCTAGATCAAAATGTTTAAATTCAGAGGGCTT 822
Db 580 CACCAAAATATTTGGAGGCTACTGGAGATCTCAAAATCAAGTGTCTCCACTGCCAGGCAT 639
QY 823 TCAGATACCTTTTGCATATCTTTGATATATAAATTCAGAGGCTTGAATTCAGAGGTGC 882
Db 640 TCAGACACTTTTGGACCTTTACCTGGACATCGCCCTGGATATCCAGGCACTCAGAGTGC 699
QY 883 AACAGGCACTTGGAGCAGTGTGTGAAGCGGCAACAGCTTGAATGGAGAAAACCTCGTACAAG 942
Db 700 AGCAAGCTTTGGAAACAGTGTGTGAAGCGGCAACAGCTTGAATGGAGAAAACCTCGTACAAG 942
QY 700 CAGCAAGCTTTGGAAACAGTGTGTGAAGCGGCAACAGCTTGAATGGAGAAAACCTCGTACAAG 942
Db 700 CAGCAAGCTTTGGAAACAGTGTGTGAAGCGGCAACAGCTTGAATGGAGAAAACCTCGTACAAG 942
QY 943 TSCAGCAAGTGTAAAGAGTGTTCAGCTTCAAGAGGTTTCACTATCCATAGATCCTCT 1002
Db 760 TGTGGTGTGTTTCTCAGAGGGCGCGCTTCCAGACGTTTAACTTTACACACTCTGCC 819
QY 1003 AATGTTCTTACATTTCTCTGAAACGTTTTCGAAATTTTACCGGTGGAAATTTGCTAAG 1062
Db 820 AAGGCTCTCATCTTGTATTGAGAGATTTCTCGATGTACAGGCAACAGATTGCCAG 879
QY 1063 GATGTGAATATACCTGAGTATCTTGATATTTGGGCAATATATGCTCAACCCAGGAGAG 1122
Db 880 AATGTGAATATCTGAGTGCCTTGCATGACGACCATACATGTCTCAGACGAAACACAGGA 939
QY 1123 CCAATGTCTAGCTTGTATGAGTGTGCTGCTCCACACTGTTTAAATTCGCAATGCTGCC 1182
Db 940 CTTCTGCTATGCTCTTATGCTGTGCTGCTCCAGCTGGTGGAGTTGTCAACAGGA 999
QY 1183 CATTAATCTCTGCTACATAAAGCTAGCAATGGCTCTGCTGATCAAAATGAATGACTCCATT 1242
Db 1000 CATTAATCTCTTATGTCAAAGCTCAAGAGGCGCAGTGGTATAAATGGATGATGCGAG 1059
QY 1243 GTATCTACAGTATATATGATCGGTACTCAGGCAACAGGCTATGCTCTTTTATATC 1302
Db 1060 GTACCGGCTCTAGCATCACTTCTGCTGAGTCAACAGGCTTACGCTCTCTTTTATATC 1119
QY 1303 AGGTCCCATGATGTGAAATATGGAGTGA 1331
Db 1120 CAGAAGAGTGAATGGGAAAGACACAGTGA 1148

RESULT 14

US-10-371-905A-25
; Sequence 25, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang

; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: US/2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 1593
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: human DUB8.1
US-10-371-905A-25

Query Match 17.8%; Score 430.2; DB 13; Length 1593;

Best Local Similarity 62.0%; Pred. No. 1.1e-127;

Matches 725; Conservative 0; Mismatches 423; Indels 21; Gaps 2;

QY 163 ATGGATGCAAGTTCTCCAGCTGGGCTGCTGCTCTTCAATGATGATGTCATATCAC 222
Db 1 ATGGGGGAGACTCACTCTACTTTGGAGGTGAGTGCAGGTTCAACCACTTTCAAATC 60
QY 223 ACATTTTCTTTAGGAACAGTACCTGCTGTGTAGTTTATTCAGATTCATCTGTACCTGAT 282
Db 61 ACATCTTCTCGGCCAATGCAAGCTTTTGTGAATCCA---GCGGACTTCTCTCCCTGAG 117
QY 283 AATCAAAACCACTCACCAAAAGATCAAGCCCTAGGTGATGGCATCGCTCTCTCCACAG 342
Db 118 AAGTCAACCACTCTCATCTGAGACCCGCTGTCGACCTCTGTGATGATTTGGCTCTGTGGCA 177
QY 343 AAGTTCTTTTCCACTCTGAGAGATTTGCTTAAGTGGCAACAAACTCATAGTTGGA 402
Db 178 AGACGCTCGCTCCAGGAGAGCTTCTCTGAGTAGCAGAGACCTGCTCGGTGGGG 237
QY 403 GCTGGGCTCAGAAATTTGGGCAATACCTGTTTGGCAATGCAGCACTGCAAGTGTTTAAAC 462
Db 238 GCTGGGCTCAGAAATATGGAAATACCTGCTACGAGAACGCTTCCCTGCAAGTGCCTGACA 297
QY 463 TACACACCACTTCTTCCAAATATCATGCTATCATATGAACACTTCCAAACATGCTATGCA 522
Db 298 TACACACTGCTTCTTCCAACTACATGCTGTCCGGGAGCACTCTCAAAACATGTCAGCGT 357
QY 523 GAAGGCTTTTGTATGATGTGTACAAATGCAAGCAATATTTACCCAGGCACTCACTATCCT 582
Db 358 CCAAGTGTGCTATGCTCTGCTACTATGCAAGTTCATCATGAGGCTTCCAGAGTCT 417
QY 583 GGGGAGCTTATTAACCAATGTTTGTCAATGAGATGCGGCTATAGCTAGGCACTC 642
Db 418 GGCCA-----TGTCATCCAGCCCTCACAGGCATTGGCTGCTGCTTC 459
QY 643 CGTTTTTGGAAACCAATAGATGCCATGAATTCCTTCAATATACACTGTTGATGATGCGAG 702
Db 460 CATAGAGGCAAGCAGTGAAGATGTCCATGAATTTCTCATGTTCACTGTGATGCTCATGAAA 519
QY 703 AAGCATGCTTGAATGCGAGCAATAAATTAGACAGACACACCCAGGCACTCTTGT 762
Db 520 AAGGCTATGCTTCCCGGCAACAGCAGGTGATGATCATCTGCAAGGACACCACTCATC 579
QY 763 TGTTCAGATTTTGGAGATACCTAAGCTTGAAGTCAAAATGTTTAAATTCAGAGGGCTT 822
Db 580 CACCAAAATATTTGGAGCTGCTGAGATCTCAAAATCAAGTGTCTCCACTGCCAGGAT 639
QY 823 TCAGATACCTTTGATTCATATCTTTGATATATAAATTCAGAGGCTTGAATGCTCAGAGTGC 882
Db 640 TCAGACACTTTTGGACCTTTACCTGGACATCGCCCTGGATATCCAGGCACTCAGAGTGC 699
QY 883 AACAGGCACTTGGAGCAGTGTGTGAAGCGGCAACAGCTTGAATGGAGAAAACCTCGTACAAG 942
Db 700 AGCAAGCTTTGGAAACAGTGTGTGAAGCGGCAACAGCTTGAATGGAGAAAACCTCGTACAAG 942
QY 943 TGCAGCAAGTGTAAAGATGTTTCCAGTTCAGAGGTTTCACTATCCATAGATCCTCT 1002

Db 760 TCGGCTTTTCTCCAGAGGGCGCGCTCCACAGTTAACTTTACACACTTCTGCC 819
Qy 1003 AATGTTCTTACACTTCTCTGAAAGCTTTTGCATAATTTTCCGGTGGAAAAATGCTAAG 1062
Db 820 AAGGTCCTCATCTTCTTGAAGAGATTCTCGATGTCGACGGAACAACAATTGCCAAG 879
Qy 1063 GATGTGAATACCTTGAGTATCTTGATATTGGCCCATATATGTCTCAACCCCAACGGAGAG 1122
Db 880 AATGTGAATATCTTGAGTGCTTGCATATGACATGACGACCATACATGTCTCAGCAGAACACAGGA 939
Qy 1123 CCAATTGTCTAGCTTCTGATGCACTGTGTGTCACACTGGTGTGTTTAAATGTCATCTGCC 1182
Db 940 CCTTGTCTATGCTCTATGCTGTGTGTGTCACGCTGGTGTGAGTGTGTACACACAGGA 999
Qy 1183 CATTAATCTCTCTCATATAAAGCTAGCAATGCCCTCTGTTATCAATGAATGACTCCATT 1242
Db 1000 CATTAATCTCTCTCATATAAAGCTCAAGAAGTCCAGTGGTATATAAATGGATGGCCGAG 1059
Qy 1243 GTATCTACAGTGATATTAGATCGGTACTCAGCCCAACAGCCATATGCTCTTTTATATC 1302
Db 1060 GTCACTGTCTGTAGCATCAATTTCTGTCTGAGTCAACAGGCGCTATGTCTCTTTTACATC 1119
Qy 1303 AGGTCCTCATGATGTGAATAATGGAGGTGA 1331
Db 1120 CAGAAGAGTGAATGGAAAGACACAGTGA 1148

RESULT 15

US-09-372-348-10
; Sequence 10, Application US/09372348A
; Publication No. US20030028005A1
; GENERAL INFORMATION:
; APPLICANT: Hedrick, Joseph A.
; APPLICANT: Wang, Luquan
; APPLICANT: Zlotnik, Albert
; APPLICANT: Murgolo, Nicholas J.
; APPLICANT: Greene, Jonathan R.
; APPLICANT: Johnston, James A.
; APPLICANT: Bazan, J. Fernando
; TITLE OF INVENTION: Mammalian Proteins; Related Reagents and Methods
; FILE REFERENCE: DX0884K(2d)
; CURRENT APPLICATION NUMBER: US/09/372,348A
; CURRENT FILING DATE: 1999-08-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1683
; TYPE: DNA
; ORGANISM: primate
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1590)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (837)
; OTHER INFORMATION: nucleotide 837 may be a or c; resulting amino acid
; OTHER INFORMATION: depends upon genetic code

US-09-372-348-10
Query Match 17.8%; Score 430.2; DB 10; Length 1683;
Best Local Similarity 62.0%; Pred. No. 1.2e-127;
Matches 725; Conservative 0; Mismatches 423; Indels 21; Gaps 2;
Qy 163 ATGGATGCAAGTCTTCCAGCTGGGGTGTGTCTTCAATGAATGATGTGTCAATCAC 222
Db 1 ATGGAGGACGACTCACTTACTTGGAGGTGAGTGGCAGTCAACCACTTTTCAAACTC 60
Qy 223 ACATTTCTTTAGGACCACTGCTGGTGTGTGTATTTATTCAGATTCACTGACTGAT 282
Db 61 ACATCTTCTCGGCCAGATGCAGCTTTTGTGTAATCCCA---CGGACTTCTCTCCCTGAG 117
Qy 283 AAATCAAAACCATCACCAAAAGGATCAAGCCCTAGGTGGCATCGCTCCTCCACAG 342

Db 118 AAGTCAACACTCTCTATCTGAGGCGCGCTGTGACCTCTGTGATGATTTGCTCCTGTGGCA 177
Qy 343 AAGTGTCTTTTCCCATCTGAGAAGATTTGTCTTAAGTGGCAACAAACTCATAGAGTTGGA 402
Db 178 AGACAGCTTGTCTCCAGGGAGAAGCTTCTCTGAGTAGCAGGAGACCTGTCTGCGGTGGGG 237
Qy 403 GCTGGGCTCCAGAATTTGGGCAATACCTGTTTGGCAATGACAGCACTGCAGTGTTTAAAC 462
Db 238 GCTGGGCTCCAGAATATGGGAAATACCTGCTACGAGAAAGCTTCCCTCAGTGGCTGCA 297
Qy 463 TACACACCACTCTTGGCAATTTACATGCTATACATGAACACTCCAAACACTGTCTATGCA 522
Db 298 TACACACCGCCCTTGGCAACTCATGCTGTCCGGGAGCACTCTCAAACTGTGAGCGT 357
Qy 523 GAAGGCTTTTGTATGATGTGTAATGAACACACATATTACCCAGGCACTCAGTAACTCT 582
Db 358 CCCAAGTCTCATGCTCTGTACTATGCAAGTCACTACATCATGGGCGCTCCACAGTCTCT 417
Qy 583 GGGACGCTTATTAACCAATGTTTGTGATCAATGAGATGGCGGTATAGCTAGGACACCTC 642
Db 418 GGTCA-----TGTCTATCCAGCCCTCACAGGCAATGGCTGCTGGCTTTC 459
Qy 643 CGTTTGGAAAACCAAGAGATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGAG 702
Db 460 CATAGAGCAAGCAGGAGAGATGCCCATGAATTTCTCATGTTCACTGTGTGATGCCATGAAA 519
Qy 703 AAAGCATCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCACCACTCTTGT 762
Db 520 AAGGCATGCTTCCCGGCCACAAGCAGGTAGATCATCTCTTAAGGACACCACTCATC 579
Qy 763 TGTGAGATATTGGAGGATACCTAAGATCTAGATCAATGTTTAAATGCAAGGCGCTT 822
Db 580 CACCAATATTGGAGGCTGCTGGAGATCTCAATCAAGTGTCTCCACTGCCACCGGAT 639
Qy 823 TCAGATACCTTTTGTATCCATATCTTGATATACTTGGAGATAAAGGCTCTCAGAGTGT 882
Db 640 CGAGACATTTTGGACCTTACTTGGACATCGCCCTGGATATCCAGGAGCTCAGAGTGT 699
Qy 883 AACAGGCAATGGAGCAGTCTTGTGAAGCCGGAACAGCTTGTATGGAGAAACTCGTACAG 942
Db 700 AAGCAAGCTTTGGAACAGTGTGTGTAAGCCGGAAGAACTCAATGGAGAGATGCCATCAT 759
Qy 943 TGCAGCAAGTGTAAAAGATGTTTCCAGCTTCAAGAGTTCACCTATCCATAGATCTCT 1002
Db 760 TGTGCTCTTGTCTCCAGAGGCGCGCCCTCCAAGACGTTAACTTTACACACTTCTGCC 819
Qy 1003 AATGTTCTTACACTTCTCTGAAACGTTTTGCAAAATTTTACGGTGGAAAAATTCCTAAG 1062
Db 820 AAGTCTCTCATCTTGTGTTTGAAGAGATCTCCGATGTCAAGGCAACAACCTTGCCAG 879
Qy 1063 GATGTGAATACCTTGAGTATCTTGAATATCGGCCATATATGTCTCAACCCCAACGGAGAG 1122
Db 880 AATGTGCAATATCTGAGTGCTTGCATGACATGACGCCATACATGTCTCAGCAGAACACAGGA 939
Qy 1123 CCAATGCTAGGCTCTGTATGCACTGCTGTGTCACACTGGTGTGTTTAAATGCACTGCTGGC 1182
Db 940 CCTCTTGTATGTCCTCTATGCTGTGTGTGTCACGCTGGGTGGAGTGTGCACAAACGGA 999
Qy 1183 CATTAATCTCTGTCATATAAAGCTAGCAATGGCCCTCTGTTATCAATGAATGACTCCATT 1242
Db 1000 CATTAATCTCTTATGTCAAGCTCAAGAAGCCAGTGTGTTAAATGGATGATGCCAG 1059
Qy 1243 GTATCTACAGTGATATTAGATCGGTACTCAGCCCAACAGCCATATGCTCTTTTATATC 1302
Db 1060 GTCAACCGCTCTAGCATCACTTCTGTCTGAGTCAACAGGCGCTAGCTCTCTTTTACATC 1119
Qy 1303 AGTCCCACTGATGTGAATAATGGAGGTGA 1331
Db 1120 CAGAAGAGTGAATGGAAAGACACAGTGA 1148

Search completed: August 17, 2004, 00:13:44

Job time : 1507 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:55:33 ; Search time 13 Seconds
(without alignments)
3104.182 Million cell updates/sec

Title: US-10-049-745-4
Perfect score: 4070
Sequence: 1 MTIVDKASSESPSAQNQP.....AAESLEPPDAASLFFPSEF 775

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4022	98.8	1198	1 UB42 HUMAN	Q9h9j4 homo sapien
2	1422.5	35.0	1121	1 UB36 HUMAN	O9p275 homo sapien
3	846.5	20.8	526	1 UB2W MOUSE	O61068 mus musculus
4	428.5	10.5	593	1 UB22 HUMAN	Q9u0t9 homo sapien
5	428	10.5	471	1 UB28 YEAST	P50102 saccharomyc
6	403	9.9	520	1 UB33 MOUSE	Q91w36 mus musculus
7	395.5	9.7	521	1 UB33 HUMAN	O9y614 homo sapien
8	392	9.6	605	1 UB22 HUMAN	O75604 homo sapien
9	385.5	9.5	357	1 UB22 CHICK	O57429 gallus gall
10	383.5	9.4	353	1 UB22 MOUSE	O88623 mus musculus
11	358.5	8.8	449	1 UB28 SCHPO	Q09738 schizosacch
12	356.5	8.8	566	1 UB21 MOUSE	Q9qz16 mus musculus
13	352.5	8.7	366	1 UB46 HUMAN	O80v95 homo sapien
14	349	8.6	565	1 UB21 HUMAN	O9uk80 homo sapien
15	343.5	8.4	792	1 UB2A YEAST	P53874 saccharomyc
16	342	8.4	355	1 UB12 HUMAN	O75317 homo sapien
17	336.5	8.3	712	1 UB44 HUMAN	Q9h0e7 homo sapien
18	332	8.2	805	1 UB25 YEAST	P39944 saccharomyc
19	324.5	8.0	1118	1 UB28 HUMAN	P40818 homo sapien
20	310	7.6	410	1 UB2X CABEL	P34547 caenorhabdi
21	307.5	7.6	977	1 UB24 HUMAN	Q9u0u5 homo sapien
22	306	7.5	898	1 UB2E DROME	Q24574 drosophila
23	304	7.5	926	1 UB29 YEAST	P32571 saccharomyc
24	295.5	7.3	585	1 UB29 SCHPO	Q9p7v9 schizosacch
25	295.5	7.3	942	1 UB33 HUMAN	O8tey7 homo sapien
26	287.5	7.1	798	1 UB10 HUMAN	O14694 homo sapien
27	279	6.9	792	1 UB10 MOUSE	P52479 mus musculus
28	275.5	6.8	1142	1 UB40 HUMAN	Q9nves homo sapien
29	270.5	6.6	2547	1 FAFX HUMAN	Q93008 h probable
30	268.5	6.6	1140	1 UB40 MOUSE	O60139 mus musculus
31	267	6.6	438	1 UB24 SCHPO	O60139 schizosacch
32	266.5	6.5	1102	1 UB27 HUMAN	Q93009 homo sapien
33	266.5	6.5	2555	1 FAFY_HUMAN	O00507 h probable

34	264.5	6.5	2559	1 FAFX_MOUSE	P70398 m probable
35	260	6.4	2778	1 FAF_DROME	P55824 drosophila
36	257	6.3	512	1 UB23 SCHPO	O94269 schizosacch
37	257	6.3	1129	1 UB2L SCHPO	Q9utt1 schizosacch
38	256	6.3	754	1 UB29 YEAST	P39967 saccharomyc
39	255.5	6.3	913	1 UB20_HUMAN	Q9y2k6 homo sapien
40	253	6.2	849	1 UB21 SCHPO	Q9usm5 schizosacch
41	251	6.2	1108	1 UB25 SCHPO	Q09879 schizosacch
42	250.5	6.2	1230	1 UB2F YEAST	P50101 saccharomyc
43	246.5	6.1	823	1 UB16 HUMAN	Q9y5t5 homo sapien
44	236.5	5.8	1042	1 UB38 HUMAN	Q8nb14 homo sapien
45	236.5	5.8	1042	1 UB38_MOUSE	Q8bw70 mus musculus

ALIGNMENTS

RESULT 1					
UB42_HUMAN					
ID	UB42_HUMAN	STANDARD;	PRT;	1198 AA.	
AC	Q9H9J4;				
DT	10-OCT-2003 (Rel. 42, Created)				
DT	10-OCT-2003 (Rel. 42, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Ubiquitin carboxyl-terminal hydrolase 42 (EC 3.1.2.15) (Ubiquitin thiolesterase 42) (Ubiquitin-specific processing protease 42) (Deubiquitinating enzyme 42) (Fragment).				
DE	USP42.				
GN	USP42.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	ISOgai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,				
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,				
RA	Nagatsuna M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,				
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,				
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,				
RA	Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,				
RA	Ninomiya K., Iwayanagi T.;				
RT	"NEDO human cDNA sequencing project."				
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.				
CC	- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =				
CC	ubiquitin + a thiol.				
CC	- SIMILARITY: Belongs to peptidase family C19.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; AK022759; BAB14232.1; -				
DR	MEROPS; C19.048; -				
DR	Genew; HGNC:20068; USP42.				
DR	InterPro; IPR001394; Peptidase_C19.				
DR	Pfam; PF00443; UCH_1.				
DR	PROSITE; PS00972; UCH_2_1; FALSE_NEG.				
DR	PROSITE; PS00973; UCH_2_2; 1.				
DR	PROSITE; PS0235; UCH_2_3; 1.				
KW	Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family.				
FT	ACT_SITE 120 120				
FT	ACT_SITE 362 362				
FT	ACT_SITE 371 371				
FT	NON_TER 1198 1198				
SQ	SEQUENCE 1198 AA; 130587 MW; CF0FDB5184AE9536 CRC64;				
Query Match 98.8%; Score 4022; DB 1; Length 1198;					
Best Local Similarity 99.6%; Pred.No.2.5e-209;					
Matches 766; Conservative 0; Mismatches 3; Indels 0; Gaps 0;					

Db 114 WERVFRVAGLHNLGNTCFNLNATQCLTVTPPLANVLLSKEHARSCHQSGFCMLCVMQNH 173
QY 163 ITOALSNPGDVLKPMFVINEMRRIARHLRGNQDEAHEFLQYTVDMOKACLINGSKILDR 222
Db 174 IVOAFANSNAIKPVSFIRDLKIAHRFRGNQDEAHEFLRYTIDAMOKACLINGSKILDR 233
QY 223 HTQATTLVQVIFGGYLSRSVKLNCCKGSDTDPYLDITLEIKAAQSVNKALQFVKPEQ 282
Db 234 QTOATTLVHQIFGGYLSRSVKCSCKVSDTDPYLDVALEIRQANIVRALFLFVADV 293
QY 283 LQGENYKSCCKKQMPVAPSKRFTIHRSSNNVLTSLKRFANFTGGKIAKOVKYPEYLDIRP 342
Db 294 LSGENAYMAKCKKVPKASKRFTIHRSSNNVLTSLKRFANFTGGKIAKOVKYPEYLDIRP 353
QY 343 YNSQNGEPIVYLYAVLHGTGNCCHAGYFCYIKASGLWYQNDSTVSTDIRSVLSQ 402
Db 354 YMSQNGDPMYGLYAVLHGTGNCCHAGYFCYIKASGLWYQNDSTVSTDIRSVLSQ 413
QY 403 QAYVLYFIRSHDVKNKGELTHPHTSPG-OSSPRPVISQSVVTKNQAPGFIGPO-LPSHM 460
Db 414 QAYVLYFIRSHDVKNKGELTHPHTSPG-OSSPRPVISQSVVTKNQAPGFIGPO-LPSHM 460
QY 461 IKNPPLHNGTGPL--KDTFSSMSPNNGSNVNRASPVNASASVQWNSVNRSSVPEHPK 518
Db 454 KKNIGNGLITSSPLTKRQDQSGTMKPHTEEI--GVPISRNGSTGLKQNGCIPPKLPS 511
QY 519 KQKITIISHKLPVRCQSQPNLHNSLENP-----TKPVPS-----SITINSVOSTNAST 571
Db 512 GS-----PSPKLSQTPHTMPTILDDPGKKVPAPQPHFSPTAQLGFTGNSN- 560
QY 572 MSVSSKVTKPIPRSESCSQPVNMGSKLNSVLV-PYGAESSDSDESKGLKENGIGT 630
Db 561 ---SSRSQORQGSWDSRDVLVSTSPKLLATATANGHLK-----GNDSAGLDR-RGSS 612
QY 631 IVSSHSPQDAEAEATPHELQEPMTLNGANSADSDPKENGAPDGASCOQPALHSE 690
Db 613 SSPEHSASSDSTKAPOTPRS-----GAALHCDSE-----TNC--STAGHSK 652
QY 691 NPPAKANGLPGLM-----PAPLSLEDKILEFRLSNKUKLGTGSDMSAPGAER 740
Db 653 TTPSGADSKTVLKSPLVSNNTTTEPASTMSPPPAK-----KLALSAKASTLWRATG--- 704
QY 741 GPPEDRADBPQCSAAESLEP-----DAAASLFP 771
Db 705 -----NDLRPPPPSSD-LTHPMKTSHPVASTWP 734
RESULT 3
UBPW MOUSE STANDARD; PRT; 526 AA.
AC Q61058;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase DUB-1 (EC 3.1.2.15) (Ubiquitin
thiolesterase DUB-1) (Ubiquitin-specific processing protease DUB-1)
DE (Deubiquitinating enzyme 1).
GN DUB1 OR DUB-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN
RP SEQUENCE FROM N.A., AND MUTAGENESIS OF CVS-60.
RX MEDLINE=96194957; PubMed=8622927;
RA Zhu Y., Carroll M., Papa F.R., Hochstrasser M., D'Andrea A.D.;
RT "DUB-1, a deubiquitinating enzyme with growth-suppressing activity.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:3275-3279(1996).
CC -!- FUNCTION: Has growth-suppressing activity, induces arrest in G1
CC phase upon controlled expression.
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- INDUCTION: By interleukin-3.

CC -!- SIMILARITY: Belongs to peptidase family C19.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U41636; AAC52532.1; -;
DR PIR; JC6133; JC6133.
DR MEROPS; C19.031;
DR MGD; MGI:107699; DUB1.
DR GO; GO:0004843; Fubiquitin-specific protease activity; IDA.
DR GO; GO:0016579; P:protein deubiquitination; IDA.
DR InterPro; IPR001394; Peptidase_C19.
DR Pfam; PF00443; UCH; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS0235; UCH_2_3; 1.
KW Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family.
FT ACT SITE 60 60
FT ACT SITE 298 298 BY SIMILARITY.
FT ACT SITE 307 307 BY SIMILARITY.
FT MUTAGEN C->S: LOSS OF ACTIVITY.
SQ SEQUENCE 526 AA; 59073 MW; 263AA7B7579694EA CRC64;
Query Match 20.8%; Score 846.5; DB 1; Length 526;
Best Local Similarity 37.1%; Pred. No. 9.6e-39;
Matches 186; Conservative 89; Mismatches 152; Indels 75; Gaps 12;
QY 76 PKDQALGDIAPP-----QKLVFPSEKICLKWQOQTHRVGAGLQNTGTCFANAALQ 127
Db 8 PEADPALSPDAPLHQAQVVEELTVNGKSLSWESQPGCGQLQNTGNSCYLNAALQ 67
QY 128 CLTYTPPLANYMLSEHSKTCBAEGFCMCTMQAHTQAL--SNPGDVTKPMFVINEMR 185
Db 68 CLTHTPPLADYMLSQEHSQTCSPGCKLCAMEALVTQSLHSHSGDVNMPKSHILTS--- 124
QY 186 IARHLRFGNQEDAHFLOVTVDMOKACLINGSKILDRHTQATTLVQVIFGGYLSRSVKC 245
Db 125 -AFHKH--QEDAHFELMFTLETMHESCLQVROKPSDESPHIDIFGGWRSQIKCL 181
QY 246 NCKGVSDTDPYLDITLEIKAAQSVNKALQFVKPEQDLGSEYKSCCKKMPVAPSKRFT 305
Db 182 LCQGTSDTYDRELDIPLDISSAQSVKQALWDTEKSEELCGDNAYYCGKCRQMPASKTLH 241
QY 306 IHRSSNVLTSLKRFANFTGGKIAKOVKYPEYLDIRPYNQNGEPIVYLYAVLHVTGF 365
Db 242 VHIAPKVLMVNLNRFSAFTGNKLDKRVSYPEFLDLKPYLSEPTGGPLPYALVAVLHDGA 301
QY 365 NCHAGHYFCYIKASGLWYQNDSTVSTDIRSVLSQQAAYVLYFIR-----SHDVKNG 418
Db 302 TSHSGHYFCVCKAGHGKWKMDTKVRCVTSVLNENAYVLYFYVQANLKQVSDIMPEG 361
QY 419 --GELTHPTH-----SPGSSPPRPVISQ-----RVVTNQ 446
Db 362 RINEVLDPEYQLKGRKKKKKSPFTDLGEPENRDKRAIKETSLGKGLVQEVNHHK 421
QY 447 AAPGF-----IGPOLPSHMLKNPPLHNGTGPLKDTPTSSS--MSSPNGNSVNRASP 495
Db 422 AGQKHGNTKLMPPQKHQ--KAGQNLNRTVEVLDLPAIDAIVIHQPRSTANWGRSDPKENQ 480
QY 496 -----VNASASVQWNSVNR 509
Db 481 PLHNADRLTSSQGVNTWQLCR 502
RESULT 4
UB22 HUMAN
ID UB22 HUMAN
AC Q9UFT9;
STANDARD; PRT; 593 AA.


```

DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase 22 (EC 3.1.2.15) (Ubiquitin
DE thiolesterase 22) (Ubiquitin-specific processing protease 22)
DE (Deubiquitinating enzyme 22) (Fragment).
GN USP22 OR KIAA1063.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:197-205(1999).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB028986; BAA83015.1; -.
CC DR MEROPS; C19.035; -.
CC DR Genew; HGNC:12621; USP22.
CC DR InterPro; IPR001394; Peptidase_C19.
CC DR InterPro; IPR001607; ZnF_UBP.
CC DR Pfam; PF00443; UCH; 1.
CC DR Pfam; PF02148; zf-UBP; 1.
CC DR PROSITE; PS00972; UCH_2_1; 1.
CC DR PROSITE; PS00973; UCH_2_2; 1.
CC DR PROSITE; PS0235; UCH_2_3; 1.
CC Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
KW NON TER
FT ACT SITE 253 BY SIMILARITY.
FT ACT SITE 539 BY SIMILARITY.
FT ACT SITE 547 BY SIMILARITY.
FT ACT SITE 547 BY SIMILARITY.
SQ SEQUENCE 593 AA; 66605 MW; 862ED0E3ED746BEF CRC64;

Query Match 10.5%; Score 428.5; DB 1; Length 593;
Best Local Similarity 30.4%; Pred. No. 3.6e-16;
Matches 104; Conservative 55; Mismatches 138; Indels 45; Gaps 6;

QY 112 GLQLNGTCFANAALQCLTYTPPLANYMLSHSKTC--HAEGFCMMCTMQ 171
DB 245 GLINLNGTCFNCVQALTHPLDLDFLSBHRCEMSPSSCLVCMSLSLFPQYSGHR 304
QY 172 DVIKPMFVINEMRRIARHLRFGNQDAHEFLQYTVDMQKACLNGSN--KLDRLHTQATLL 229
DB 305 SPHIPPYKLLHLVWTHARHLAGVEQQAHEFLIALDLVLRHCKGDDNGKANNPNHCNCI 364
QY 230 VCQIFGGYLRVRKCLNCKGVSDFPDYDITLE-----I 264
DB 365 IDQITFGTLQSDVTCQVCHGVSTTIDPFDWISLDLPFGSSTPFWPLSPSGEKNVNGSHV 424
QY 265 KAAGSVNKALEQFVKPEQLDGENSVKCSKCKKWVPASKRFTTHRSSNVLTLSLKRFANFT 324
DB 425 SGTYTLTCLRRFTPEHLGSSAKIKCSGCHSYQESTKQLTWKLLPIVACFLKXFEHSA 484
QY 325 --GGKIADKVYPEYLDIRPYW-----SQPNGE-----PIYVLYAVLHTGNC 367
DB 485 KLRKKITTVSGFPLELDTFPWASSKESRMNGQYQQPTDSLNDNDNKYSLEAVVNHQG-TL 543

```

```

QY 368 HAGHYFCYIKASGLNGLYQMNDSTVSTDIRSVLSQQAVVLFY 409
DB 544 ESGHYTFIRQHKDQWFKCDALITKASIKDVLDSGLLYLFY 585

RESULT 5
UBP8_YEAST
ID UBEP8_YEAST STANDARD; PRT; 471 AA.
AC P50102;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase 8 (EC 3.1.2.15) (Ubiquitin
DE thiolesterase 8) (Ubiquitin-specific processing protease 8)
DE (Deubiquitinating enzyme 8).
GN UBEP8 OR YMR223W OR YMR959.05.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313268; PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Lye G., Moulte S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII.";
RL Nature 387:90-93(1997).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z49939; GAA90194.; -.
CC DR FIR; S57591; S57591.
CC DR GenOnline; 142898; -.
CC DR MEROPS; C19.0PW; -.
CC DR SGG; S0004836; UBP8.
CC DR GO; GO:0000124; C:SAGA complex; IDA.
CC DR InterPro; IPR001394; Peptidase_C19.
CC DR InterPro; IPR001607; ZnF_UBP.
CC DR Pfam; PF00443; UCH; 1.
CC DR Pfam; PF02148; zf-UBP; 1.
CC DR SMART; SMO0290; ZnF_UBP; 1.
CC DR PROSITE; PS00972; UCH_2_1; 1.
CC DR PROSITE; PS00973; UCH_2_2; 1.
CC DR PROSITE; PS0235; UCH_2_3; 1.
CC Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
KW ACT SITE 146 BY SIMILARITY.
FT ACT SITE 146 BY SIMILARITY.
FT ACT SITE 419 BY SIMILARITY.
FT ACT SITE 427 BY SIMILARITY.
SQ SEQUENCE 471 AA; 53633 MW; BC632F12FBD0F73C CRC64;

Query Match 10.5%; Score 428; DB 1; Length 471;
Best Local Similarity 31.4%; Pred. No. 2.9e-16;
Matches 104; Conservative 68; Mismatches 125; Indels 34; Gaps 11;

QY 111 AGLNLGNTCFANAALQCLTYTPPLANYMLSHSKTC--HAEGFCMMCTMQ----- 160
DB 137 SGLNKGSTCFMSSILQCLIHNPYFIRHSMQIHSNCKVSRSPDKCFSCALDKVHLYLG 196
QY 161 -AHITQALSNPDGV IKP---MFVINEMRIARHLRFGNQDAHEFLQYTVDMQKAC---- 213

```


Db 197 ALNTKQASSSTSTNRQTGFYLLTCAWKINQNLQAGYSQDAHEFWQFIHQHVSULD 256
 QY 214 LNSGNKLDL--HTQATTLVCQIPGGYLRGRVCLNCKGVS-DTFDPLDITLLEIKAAQSV 270
 Db 257 LPNAKEVSRRANKQCEIVHTVFEGLSSIVCPGQNNKSTTIDPFLDLSLDIKDKKL 316
 QY 271 NKALBOFVKPEOLDGNSYKCKCKKQWVASKFTTHRSSNVLTSLKFAFNTGG--K 327
 Db 317 YECLOSDFHKEQKQDFN-YHCGECNSTQDAIKQLGHKLPLSVLQKRFPELLNGSNRK 375
 QY 328 IAKDVKYPEYLDIRPMS-----OPNGE--PIVVVLYAVLVHTGFNCHAGHYFCYIKA 378
 Db 376 LDDFIEFFYLLNMKNVCSYKXKDHSENGKVPDIIVELIGIVSHKG-TYNEGHYIAFCKI 434
 QY 379 SGNLWQNMDSIVSTSDIRSVLSQQAYVLVY 409
 Db 435 SGGQWFKFNDMSVSSISQBEVLKEQAYLLFY 465

RESULT 6

UBP3 MOUSE STANDARD; PRT; 520 AA.
 AC Q91W36;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ubiquitin carboxyl-terminal hydrolase 3 (EC 3.1.2.15) (Ubiquitin
 thiolesterase 3) (Ubiquitin-specific processing protease 3)
 DE (Deubiquitinating enzyme 3).
 GN USP3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1] _SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6J; TISSUE=Forelimb;
 RC MEDLINE=22354683; PubMed=12466851;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
 RA Baldarelli L.M., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konegaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Sempke C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wysshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary gland;
 RA MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

Klausner R.D., Collins P.S., Wagner L., Shemmen C.M., Schuler G.D.,
 Altschul S.F., Zebberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Brownstein M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,
 Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -!- SIMILARITY: Belongs to peptidase family C19.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AK031143; BAC27276.1; -;
 DR EMBL; BC017156; AAH17156.1; -;
 DR MEROPS; C19.026; -;
 DR MGD; MGI:2152450; Usp3.
 DR InterPro; IPR001394; Peptidase_C19.
 DR InterPro; IPR001607; Znf_UBP.
 DR Pfam; PF00443; UCH; 1.
 DR Pfam; PF02148; zf-UBP; 1.
 DR PROSITE; PS00972; UCH 2_1; 1.
 DR PROSITE; PS00973; UCH 2_2; 1.
 DR PROSITE; PS02935; UCH 2_3; 1.
 KW Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
 FT ACT SITE 168 168 BY SIMILARITY.
 FT ACT SITE 462 462 BY SIMILARITY.
 FT ACT SITE 471 471 BY SIMILARITY.
 SQ SEQUENCE 520 AA; 58668 MW; E160B48CA47483D CRC64;
 Query Match 9.9%; Score 403; DB 1; Length 520;
 Best Local Similarity 29.9%; Pred. No. 7.3e-15;
 Matches 115; Conservative 51; Mismatches 117; Indels 102; Gaps 13;
 QY 112 GLQNLGNTCFANALQCLT-----YTPPL-----ANYMLIS 141
 Db 160 GLRLNLTGCFNAILQSLNIFQCYFKELPVELRNGKTAGRTYHTRSGDSNVSLV 219
 QY 142 HEHSTKTHAEGFCMCTQWAHITQALNSPGDVIKPMFVINEMERIARHLRFGNQDAHEF 201
 Db 220 EEFKRT-----LCALWQGSQTAFS-----PESLFYVWVKIMPFRGYQQDAHEF 264
 QY 202 LOY-----TVDMOKACLNGSKNLDLDRHTQATLYVCOIFG 235
 Db 265 MYRLDLHLHLELQGGFNGVSRSAIQENSTLSASNKCCING-----ASTVTAIFG 315
 QY 236 GYIIRSRVKCLNCKGVSDFDPYLDITLRI-----KAAQSVNK-----ALEQFVKPE 281
 Db 316 GLQLQNEVCLICGTESRKFDPFLDLSLQIPQFRSKRNQNGPVCSLRDCIRSFDTLE 375
 QY 282 QLDGNSYKCKCKKQWVASKRFTTHRSSNVLTSLKRF--ANFTGGIAKDVKYP-EYL 338
 Db 376 ELDETELYMCHCKKKQKSTKKFWIQLPKALCLHLKRFHWTAYLRNKNVDVTVQPLRGL 435
 QY 339 DIRPYMSQPNG---EPTIVVLYAVLVHTGFNCHAGHYFCYIKASNGLWYQMDSIVSTSD 395

```

Db 436 DMKYLLEPENSQDSCLYDLAAVVHGGVSGHYTAY-AVHEGRWFHENDSTVITVD 494
QY 396 IRSVLQQAYVLFYIRSHDVKNQGE 420
Db 495 EETVGRKAKAYILFYV-BRQARAGAE 518

RESULT 7
UBP3 HUMAN
ID _UBP3 HUMAN STANDARD; PRT; 521 AA.
AC Q9Y6I4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Ubiquitin carboxyl-terminal hydrolase 3 (EC 3.1.2.15) (Ubiquitin
DE thiolesterase 3) (Ubiquitin-specific processing protease 3)
DE (Deubiquitinating enzyme 3).
GN USP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=99410422; PubMed=10480896;
RA Sloper-Mould K.E., Eyre H.J., Wang X.-W., Sutherland G.R.,
RA Baker R.T.;
RT "Characterization and chromosomal localization of USP3, a novel human
RT ubiquitin-specific protease."
RL J. Biol. Chem. 274:26878-26894 (1999).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- TISSUE SPECIFICITY: Expressed in all tissues examined, with
CC strongest expression in pancreas.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF073344; AAD42992.1; -
DR MEROPS; C19.026; -.
DR Genew; HGNC:12626; USP3.
DR MIM; 604728; -.
DR GO; GO:0004843; F:ubiquitin-specific protease activity; TAS.
DR InterPro; IPR001394; Peptidase_C19.
DR InterPro; IPR001607; ZnF_UBP.
DR Pfam; PF00443; UCH; 1.
DR Pfam; PF02148; zf-UBP; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS0235; UCH_2_3; 1.
DR Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
KW ACT SITE 168 BY SIMILARITY.
FT ACT SITE 463 BY SIMILARITY.
FT ACT SITE 472 BY SIMILARITY.
SQ SEQUENCE 521 AA; 59097 MW; EF5CF39BA5482D67 CRC64;

Query Match 9.7%; Score 395.5; DB 1; Length 521;
Best Local Similarity 30.3%; Pred. No. 1.8e-14;
Matches 117; Conservative 48; Mismatches 118; Indels 103; Gaps 14;

QY 112 GLQNLGNTCFANALQCLIT-----YTPPLA-----NYWLS 141
Db 160 GLNLGNTCFNALLQSLSNIEQCCYFKELPAVELNGKTAGRTVHTRSGQDNVSLV 219
QY 142 HEHSKTHAEGFCWGMCTWQAHTQALNPGDVIKPMFVINEMRRIARHLRFGNQEDAHEF 201

```

```

Db 220 EEFKRT-----ICALWQGSOTAFS-----PESLFFVVKIMPNFRGQQOQDAHEF 264
QY 202 -----LOYTVDAQ-----KACLNGSNKLDHRHTQATTLVCOIF 234
Db 265 NALPFGPPTLGNFRANSTVFPAQOFCRRILLCLQVKNCCING-----ASTVVTAF 315
QY 235 GGYLSRSVKCLNCKGVSDDTFDPYLDITLEI-----KAAQSVNK-----ALEQFVKP 280
Db 316 GGIQLQNEVNCLICTESRKDFDPLDISLDIPSPQSRKSKNQENGPVCSLDCLRSFTDL 375
QY 281 EQLDGENSYKSKCKKMPASKRFTTHRSSNVLTSLKRF--ANFTGGKIAKDVKYP-EY 337
Db 376 EELDELYMCHCKKKQKSTKKFWIQKLPKVLCLHLKRFHTAYLRKNKVDVTFEFLRG 435
QY 338 LDRPYMSQPNQ---EPVVVLYAVLVHTGFCNCHAGHYFCYIKASNLGLWQMNDISYVS 394
Db 436 LDMKWYLLEPENSQDSCLYDLAAVVHGGVSGHYTAY-AVHEGRWFHENDSTVITVD 494
QY 395 DIRSVLSQQAYVLFYIRSHDVKNQGE 420
Db 495 DEETVGRKAKAYILFYV-EHQAKAGSD 519

RESULT 8
UBP2 HUMAN
ID _UBP2 HUMAN STANDARD; PRT; 605 AA.
AC Q756Q4; Q9BQ21;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase 2 (EC 3.1.2.15) (Ubiquitin
DE thiolesterase 2) (Ubiquitin-specific processing protease 2)
DE (Deubiquitinating enzyme 2) (41 kDa ubiquitin-specific protease).
GN USP2 OR UBP41.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. (ISOFORM 2).
RT TISSUE=Placenta;
RC Gong L., Yeh E.T.H.;
RT "Cloning and expression of the human and mouse UBP41."
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Lymph;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg E., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.E., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.N., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;

```



```

QY 316 SLKRF--ANFTGKIAKVYP-EYLDIRPYMQNGEPIVYVLYAVLHTGFNCAGHY 372
D 253 HLKRFSEARISKLTTFFNPLKLDLREFASQ-SCNHAAYNLYAVSNHSG-TTMGGHY 310
QY 373 FCYIKAS-NGLWYQNDISVSTDIRSVLSQAAVLFYIRSHDVKNKGELTHPT 425
D 311 TAYCKSPISSEHNSFNDVRTPMSSSHVRSDDAYLLFY-----ELASPS 354

RESULT 10
UBP2_MOUSE
ID UB2_MOUSE STANDARD; PRT; 353 AA.
AC O88623;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Ubiquitin carboxyl-terminal hydrolase 2 (EC 3.1.2.15) (Ubiquitin
DE thiolesterase 2) (Ubiquitin-specific processing protease 2)
DE (Deubiquitinating enzyme 2) (41 kDa ubiquitin-specific protease).
GN USP2 OR UBPA41.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Gong L., Yeh E.T.H.;
RT "Cloning and expression of the human and mouse UBPA41."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF079565; AAC28393.1; --
CC MEROPS; C19.013; --
CC MGD; MGI:1858178; Usp2.
CC InterPro; IPR001394; Peptidase_C19.
CC Pfam; PF00443; UCH_1.
CC PROSITE; PS00972; UCH_2_1; 1.
CC PROSITE; PS00973; UCH_2_2; 1.
CC PROSITE; PS0235; UCH_2_3; 1.
CC Ub1 conjugation pathway; Hydrolase, Thiol protease; Multigene family.
FT ACT_SITE 24 24 BY SIMILARITY.
FT ACT_SITE 297 297 BY SIMILARITY.
FT ACT_SITE 305 305 BY SIMILARITY.
SQ SEQUENCE 353 AA; 40581 MW; 4FPB39A225FE8F11 CRC64;

Query Match 9.4%; Score 383.5; DB 1; Length 353;
Best Local Similarity 30.3%; Pred. No. 5e-14;
Matches 104; Conservative 55; Mismatches 127; Indels 57; Gaps 12;

QY 111 AGLNLGNTCFANAALQCLTYTPPLANYMLSH-----EHSKTHA---EGFMMCTMQA 161
D 15 AGLNLGNTCFNNSILOCLSNTRFLDYCLQRLYMRDLGHTSSANTALMEFP-----A 67
QY 162 HITQAL--SNPGDVTKMPFVINEHRIARHFRNQEDAEHFLQYTVDMQACLINGSNK 219
D 68 KLIQTIWTSPPNDVVSPEFKTIQRYAPRFGYNGQDAEFLRFLDLGLH----NEVNR 123
QY 220 LDRHTQAT-----TLVCQIFGGYILSRVKLCKNGKGVSDTF 254
D 124 VAARPKASPETLDLPDEBKGRMKRYLEREDSRIGDLFVGOLKSSLTCTCGYGVSTVF 183
QY 255 DFLVDITLBIK---AAQSVNKALEQFVKPEQUDGENSYKSKCKKMWVPASKRFTIHRSS 310

```

```

D 184 DPFWDLSLPIAKRGYPVETLMDQMLFTKEDILDGDEKPTCCCRARKRCKIKFSVQRF 243
QY 311 NVUTLSIKRF--ANFTGKIAKVYP-EYLDIRPYMQNGEPIVYVLYAVLHTGFNC 367
D 244 KILVLHLKRFSEARISKLTTFFNPLKLDLREFASE-NTNHAAYNLYAVSNHSG-TT 301
QY 368 HAGHYFCYIKAS-NGLWYQNDISVSTDIRSVLSQAAVLFY 409
D 302 MGGHYTAYCRSPVTGEWHTFNDSSVTPMSSSQVRTSDAYLLFY 344

RESULT 11
UBP8_SCHPO
ID UB8_SCHPO STANDARD; PRT; 449 AA.
AC Q09738;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE Ubiquitin carboxyl-terminal hydrolase 8 (EC 3.1.2.15)
DE (Ubiquitin thiolesterase 8) (Ubiquitin-specific processing protease 8)
DE (Deubiquitinating enzyme 8).
GN UB8 OR SPAC13A11.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Borzym K., Langer I., Back A., Leinach H., Wambutt R., Purnelle B.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z54096; CAA90805.1; --
CC FIR; T37611; T37611.
CC MEROPS; C19.UPW; --

```

RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavita T.,
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA	Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA	Gustingich S., Hall D., Hofmann J., Hume D.A., Kamiya M., Lee N.H.,
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA	Hayashizaki Y.;
RT	"Functional annotation of a full-length mouse cDNA collection.";
RL	Nature 409:685-690(2001).
CC	-!- FUNCTION: Capable of removing ubiquitin from ubiquitinated
CC	proteins; also capable of removing NEDD8 from NEDD8 conjugates but
CC	has no effect on Sentrin-1 conjugates (By similarity).
CC	-!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2O) =
CC	ubiquitin + a thiol.
CC	-!- SIMILARITY: Belongs to peptidase family C19.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; AF177759; AAD54322.1; --
DR	EMBL; AK011148; BAB27431.1; --
DR	MEROPS; C19.034; --
DR	MGI; MGI:1353665; Usp21.
DR	InterPro; IPR001394; Peptidase_C19.
DR	Pfam; PF00443; UCH; 1.
DR	PROSITE; PS00972; UCH_2_1; 1.
DR	PROSITE; PS00973; UCH_2_2; 1.
DR	PROSITE; PS02335; UCH_2_3; 1.
DR	Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
KW	ACT_SITE 221 221 BY SIMILARITY.
FT	ACT_SITE 511 511 BY SIMILARITY.
FT	ACT_SITE 519 519 BY SIMILARITY.
FT	CONFLICT 220 220 T -> TLQP (IN REF. 2).
FT	CONFLICT 338 338 R -> C (IN REF. 2).
SQ	SEQUENCE 566 AA; 62672 MW; F2E1828E9BDB1AFD CRC64;
Query Match	8.8%; Score 356.5; DB 1; Length 566;
Best Local Similarity	26.4%; Pred. No. 2.6e-12;
Matches 122; Conservative 68; Mismatches 181; Indels 91; Gaps 16;	
QY	23 SEAVSPGDM-----DAGSAGWAGVSLNDVSNHLSILGPVPGVAVYSSSSVDPKSK 73
Db	111 SKSVSGDLRPMGIALGHRGACELGALS-----RLALRPEPTLRSTSRLLGGF 163
QY	74 PSP-----QKQALGDG-----IAPQKVLFPSEKILCKWQOTHRVA---GLQNLGNT 119
Db	164 PGEPPTLLSIRTPPTSHGSHFMISARPEPFYSDDKMA---HHTLLGSHGVGLRLNGT 220
QY	120 CFANAAQLCLTTPPLANYMLSHEHSKTCCHAEFCMMCTMQ-AHITQALSNGP--DVIKP 176
Db	221 CFNLAVLQCLSTRPLRDLCLRRDFQEVPGGGAQELTEAFADVTGALMHPDSCAVNP 280
QY	177 MFVINMERIARHLRFGNQEDAHEFTLQYTVDM----- 209
Db	281 TRFAVFQKVPVFSFGYSQDAQEFLKLLMERLHLLEINRRGRAPILASGVPSPRRG 340
QY	210 -----QKACLSNGKLDRHTQATTLVCOIFGVIILSRVKLCNCKGVSTFDP 256
Db	341 GGALEHEPELSDDDRLANLMWKYLRSE---DSKIVDLFVGOLKSLCKCOACGVRSTTFE 397

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
QC the European Bioinformatics Institute. There are no restrictions on i

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
QC the European Bioinformatics Institute. There are no restrictions on i

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
QC the European Bioinformatics Institute. There are no restrictions on i

QY	23	SEAVSPGDM-----DAGSASWCAVSSLNDVSNHILSLGVPVCAVYVSSSSVDPKSK	73
Db	111	SKSVSSGDLRPMWIALGHRGTGELGAALS-----RLALRPEPTLRRSTSL-----	157
QY	74	PSPOKDAQDGIAPP-----QKLFPSKILCKWQQTHRVG	110
Db	158	-----RELGGFPGPPTLFSIRTEPPASHGSHMILSARSSEFFYDDKVA---HHTLLIG	208
QY	111	A---GLONLGNCTCPANAALQCLTVPPLANYMLSHEHSKTCCHAEFCMMCTWQ-AHITQA	166
Db	209	SGHYGLRLNIGNTCFINAVLQCLSTRPLRDFCLRRDFRQEVPGGGRQAQELTEAFADVGA	268
QY	167	LSNPG--DVLKPMFVINEMRIARHLRFGNQDAHEFLQYTVDM-----	209
Db	269	LWHPDSCAVNPTFRVQFQVPSFGSYQODAQEFLKLMLERLHLEINRRGRAPPIL	328
QY	210	-----QKACLSNGKLDRHTQATTLVCQIFGGYLRSRVKCL	245
Db	329	ANGVPVPPRRGGALLBEPCLSDDDRANLWKRYLERE---DSKLVDFLFGQLKSLCKQ	385
QY	246	NCKGVSUTFPYDLITLTK-----AAQSVNKALEQFVKPEQLDGENSYKSKCKQWMP	299
Db	386	ACGYRSTTFEVCDSLPIPKKGAGGVKVSURDCFNLFKEEELSENAPVCDRCRQKTR	445
QY	300	ASKRETHRSSNVLTLISKPFANFTGG--KIAKQVKYP-EVLDTRPYMSQPNGEPIVYL	356
Db	446	STKKLVQRPFRILVLHNLFRSASRGSIKKSSVGDFPLQRLSLGDFASDKAGSP-VYQL	504
QY	357	YAVLVHTGFNCHAGHYFCYTKASNLGMVQMNDSIVSTSDIRSVLSQQAYVLFY	409
Db	505	YALCNHSG-SVHYGHYIALCRQQTG-NHVYNDSRVSPSENVQVASSGIVLFY	555
RESULT 15			
UBPA	YEAST	STANDARD;	PRT; 792 AA.
AC	PS3874;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DE	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Ubiquitin carboxyl-terminal hydrolase 10 (BC 3.1.2.15) (Ubiquitin		
DE	thiolesterase 10) (Ubiquitin-specific processing protease 10)		
DE	(Deubiquitinating enzyme 10) (Disruptor of telomere silencing protein		
DE	4).		
GN	UBP10 OR DOT4 OR YNL186W OR N1619.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
OX	NCBI_TaxID=4932;		
RN	[1]_TaxID=4932;		
RP	SEQUENCE FROM N.A.		
RL	Obermaier B., Piravandi E., Rinke M., Domdey H.;		
RL	Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.		
RP	[2]		
RP	FUNCTION.		
RX	MEDLINE=98429517; PubMed=9755194;		
RA	Singer M.S., Mahana A., Wolf A.J., Meisinger L.L., Peterson S.E.,		
RA	Goggin C., Mahowald M., Gottschling D.E.;		
RT	"Identification of high-copy disruptors of telomeric silencing in		
RT	Saccharomyces cerevisiae.";		
RL	Genetics 150:613-632(1998).		
CC	-!- FUNCTION: Has a role in telomere and HM loci silencing, which is		
CC	the repression of chromatin structure which leads to a stop in the		
CC	transcription of nearby genes.		
CC	-!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =		
CC	ubiquitin + a thiol.		
CC	-!- SIMILARITY: Belongs to peptidase family C19.		

CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		

Wed Aug 18 13:53:06 2004

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Z71462; CAA96080.1; -
PIR; S63141; S63141.
GermOnline; 143192; -
MEROPS; C19.UPW; -
SGD; S0005130; UBP10.
GO; GO:0005634; C:nucleus; IDA.
GO; GO:0004843; F:ubiquitin-specific protease activity; IDA.
InterPro; IPR001394; Peptidase_C19.
Pfam; PF00443; UCH; 1.
PROSITE; PS00972; UCH_2_1; FALSE NEG.
PROSITE; PS00973; UCH_2_2; 1.
PROSITE; PS0235; UCH_2_3; 1.
Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family;
Telomere.
DOMAIN 127 130 POLY-SER.
DOMAIN 138 144 POLY-GLU.
DOMAIN 252 255 POLY-GLU.
DOMAIN 302 309 POLY-GLU.
ACT_SITE 371 371 BY SIMILARITY.
DOMAIN 556 562 POLY-SER.
DOMAIN 682 691 BY SIMILARITY.
ACT_SITE 691 691 BY SIMILARITY.
ACT_SITE 691 691 BY SIMILARITY.
SEQUENCE 792 AA; 88517 MW; 16CB28FD4F989874 CRC64;

Query Match 8.4%; Score 343.5; DB 1; Length 792;
Best Local Similarity 22.6%; Pred. No. 2e-11;
Matches 115; Conservative 78; Mismatches 174; Indels 141; Gaps 17;
QY 96 SEKICLKW--QOHRVGAGLQNLGNTCFANALQCLTYTPPLANYML-----SHEHSKTC 148
Db 345 STRIVKNGDKFTNLKPRGLNHGVTCTYNAAVQAMLHPSIQHYLPDILMGKYDSTISK 404
QY 149 HAEGFCMCTMQAHITQALSNP-----GDVTKPMFVNEMERIARHFRGNQDEAHEFLQ 203
Db 405 NSVSYYTLAETSKQWMLPVSKNPRKNVGSASYINPKHLISRLDDINCMSEWQEDSHEYFM 464
QY 204 YTDVAMOKACLNGSKLDRHTQATTLVCQIFGGYLRSRVKCLNCKGVSDTDPDYLDITLE 263
Db 465 SLMSRLQEDSVPKGKHL-----IESIIVDIFGGLLKQIVTCKSCGISKTEQPFYDLSLH 519
QY 264 IKAQAQ----- 268
Db 520 LKGGKLDPNSDLSDSINGTSATTSTTTSNATKPSLSSSSSVNLNNGSPFAAADLS 579
QY 269 -----SVNKALEQVKPE--QLDGE-NSYKCKCKKMPASKRFTIHRSSNVLTSLKRF 320
Db 580 ANRRFSIEKSIKDFNPFLIKVDKEQKGVCEKCHKTNAVKHSILPAPETLIVHLKXF 639
QY 321 ANFTG--GKIAKDVKPEYLDIRPYMSQNGEPIVYVLYAVLVHTGFNCHAGHYFCYIK 377
Db 640 -RFGTSSSKMKQAVSPMFLDLEY-CESKELPKYQLLSVVHGRSLSSGHYIAHCK 697
QY 378 ASGLWQMNDISVSTDIRVLQ-QAYVLFYIRSHDVKNKGELTHETHPGQSSRPV 436
Db 698 QPDGWSATYDDYINIIISERDVLKEPNAYLLYLR-----LT-----PKSVPLPL 742
QY 437 ISQRVVTNKQAAPGFIGQLPSHMIKNPPLHNGTGKPLDTPSSMSSPNGSNVNRASPV 496
Db 743 AKSAMAT-----GNVTSKSKQEQAVNEFN-----NRPLKI 772
QY 497 NASA-SVQNVSWNRSSVIEPHPKQKIT 523
Db 773 NSKKNRKKKKCN-----KKKFT 791

Search completed: August 10, 2004, 19:58:53
Job time : 15 secs

XX OS Homo sapiens.
XX PN WO200110903-A2.
XX PD 15-FEB-2001.
XX PF 09-AUG-2000; 2000WO-US021878.
XX PR 09-AUG-1999; 99US-0147986P.
XX PT 21-OCT-1999; 99US-0160807P.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Yue H, Lal P, Tang YT, Bandman O, Baughn MR, Azimzai Y, Lu DAM,
XX PI Yang J;
XX DR N-PSDB; AAF81717.
XX XX New protease (inhibitors) useful for diagnosis and treatment of
PT autoimmune/inflammatory disorders such as acquired immunodeficiency
PT syndrome, Cushing's disease, Addison's disease and cell proliferative
PT disorders such as cancer.
XX PS Claim 1; Page 95-96; 134pp; English.
XX CC AAF81714 to AAF81740 encode the human proteases and protease inhibitors
CC (PPIMs) given in AAB74668 to AAB74694. The PPIMs can have activities such
CC as: anti-human immunodeficiency virus (HIV); antidiabetic; antithyroid;
CC immunostimulant; immunomodulator; anti-inflammatory; immunosuppressive;
CC nephrotropic; antitumor; thyromimetic; cytostatic; antibacterial;
CC fungicide; proteoacid; antiarteriosclerotic; antiatherosclerotic;
CC virucide; antiparasitic; and hepatotropic. PPIM polynucleotide and
CC protein sequences can be used in the diagnosis, treatment and prevention
CC of autoimmune/inflammatory disorders such as AIDS, DiGeorge's syndrome,
CC severe combined immunodeficiency disease (SCID), Chediak-Higashi
CC syndrome, Cushing's disease, Addison's disease, autoimmune thyroiditis,
CC Crohn's disease, diabetes mellitus, Good pasture's syndrome, gout,
CC Grave's disease, Hashimoto's thyroiditis, Sjogren's syndrome, Werner's
CC syndrome, viral, bacterial, fungal, parasitic, protozoal, and helminthic
CC infections and cell proliferative disorder such as arteriosclerosis,
CC atherosclerosis, cirrhosis, hepatitis, psoriasis and cancer. PPIM
CC polynucleotide sequences can be used in somatic or germline gene therapy
CC and in diagnosis of diseases. They can also be used in generating
CC hybridisation probes useful in mapping the naturally occurring genomic
CC sequences and in molecular biology techniques
XX Sequence 775 AA;
Query Match 100.0%; Score 775; DB 4; Length 775;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 775; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTIVDKASFDSPAYONQPGSSSEAVSPGDMDSAGSAGVSSLDVSNHTLSLGPVPGA 60
DB 1 MTIVDKASFDSPAYONQPGSSSEAVSPGDMDSAGSAGVSSLDVSNHTLSLGPVPGA 60
QY 61 VYSSSSVPDKSPQKQDQALGDGTPAPKQVLPFPSEKICLKWQTHRVGAGLQNGTC 120
DB 61 VYSSSSVPDKSPQKQDQALGDGTPAPKQVLPFPSEKICLKWQTHRVGAGLQNGTC 120
QY 121 FANAALQCLTYTPPLANYMLSHHSKTCCHAEFCMCTQAHITQALSNGPQVFKPMFVI 180
DB 121 FANAALQCLTYTPPLANYMLSHHSKTCCHAEFCMCTQAHITQALSNGPQVFKPMFVI 180
QY 181 NEMRIARHLRFGNEDAEHFLQYTVDMQKACLSNKLDRHTQATTLVCCQIFGGLRS 240
DB 181 NEMRIARHLRFGNEDAEHFLQYTVDMQKACLSNKLDRHTQATTLVCCQIFGGLRS 240
QY 241 RVKCLNCKGVSTFDPLDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKCSKCKWVPA 300
DB 241 RVKCLNCKGVSTFDPLDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKCSKCKWVPA 300

QY 301 SKRFTIHRSSNVLTLSLKRFPANFTGGKIAKDVKYPEYLDIRPYMSQNGEPIVYVLYAVL 360
DB 301 SKRFTIHRSSNVLTLSLKRFPANFTGGKIAKDVKYPEYLDIRPYMSQNGEPIVYVLYAVL 360
QY 361 VHTGFCHAGHYFCYIKASNGLWYQWDSIVSTSDIRSVLSQAYVLFYIRSHDVXNGGE 420
DB 361 VHTGFCHAGHYFCYIKASNGLWYQWDSIVSTSDIRSVLSQAYVLFYIRSHDVXNGGE 420
QY 421 LTHPTSPGQSSPRPVISORVVTNKQAAFGTGPOLPSHMIKNPPHNLNGTGPKLKDTPSSS 480
DB 421 LTHPTSPGQSSPRPVISORVVTNKQAAFGTGPOLPSHMIKNPPHNLNGTGPKLKDTPSSS 480
QY 481 MSSPENGSSVNRASPVNASVQWNSVNRSSVPIPEHPKQKQITISIHNLKLPVRQCSQPN 540
DB 481 MSSPENGSSVNRASPVNASVQWNSVNRSSVPIPEHPKQKQITISIHNLKLPVRQCSQPN 540
QY 541 LHSNLENPTKVPSSITINSVQSTSNASTMSVSSKVTPIPRSSCSQPVNMGSKLN 600
DB 541 LHSNLENPTKVPSSITINSVQSTSNASTMSVSSKVTPIPRSSCSQPVNMGSKLN 600
QY 601 SSVLPVGAESSEDSDEESKGLKENGIGTIVSSHSPQDAEDEEATPHELQEPMTLNGA 660
DB 601 SSVLPVGAESSEDSDEESKGLKENGIGTIVSSHSPQDAEDEEATPHELQEPMTLNGA 660
QY 661 NSADSDSDPKENGLAPDGASCQOPALHSENPAKANGLPKGLMPAPLLSLPEDKILETF 720
DB 661 NSADSDSDPKENGLAPDGASCQOPALHSENPAKANGLPKGLMPAPLLSLPEDKILETF 720
QY 721 RLSNKLKGTDEMSAPGAERGPEDRDPAEPQGSPPAAESLEEDPAAASLFPFSESG 775
DB 721 RLSNKLKGTDEMSAPGAERGPEDRDPAEPQGSPPAAESLEEDPAAASLFPFSESG 775
RESULT 2
AAU98889
ID AAU98889 standard; protein; 1251 AA.
XX AAU98889;
AC AAU98889;
XX 26-AUG-2002 (first entry)
DE Human protease PRTS7.
KW Protease; human; gastro.intestinal disorder; gastritis; atherosclerosis;
KW ulcerative colitis; Reye's syndrome; cardiovascular disorder; enzyme;
KW hypertension; myocardia... infarction; autoimmune disease; AIDS; PRIS;
KW inflammatory disorder; acquired immunodeficiency syndrome; asthma;
KW leukaemia; developmental disorder; Cushing's syndrome; impotence;
KW epithelial disorder; dermatitis; scabies; eczema; neurological disorder;
KW Parkinson's disease; dementia; Alzheimer's disease; infertility;
KW Huntington's disease; multiple sclerosis; reproductive disorder.
XX Homo sapiens.
XX WO200238744-A2.
XX 16-MAY-2002.
XX 18-OCT-2001; 2001WO-US051034.
XX 18-OCT-2000; 2000US-0241573P.
XX 25-OCT-2000; 2000US-0243643P.
XX 02-NOV-2000; 2000US-0245256P.
XX 13-NOV-2000; 2000US-0248395P.
XX 16-NOV-2000; 2000US-0249826P.
XX 20-NOV-2000; 2000US-0252303P.
XX 01-DEC-2000; 2000US-0250981P.
XX (INCY-) INCYTE GENOMICS INC.
XX Lee EA, Hafalia AJA, Yue H, Lal PG, Yao MG, Lu Y, Walia NK;
PI

Warren BA, Lu DM, Baughn MR, Delegeans AM, Burford N, Borowsky ML;
PI Lee S, Xu Y, Griffin JA, Kallick DA, Gandhi AR, Arvizu C, Ison CH;
PI Tang YT, Azimzai Y, Elliott VS, Swarnakar A, Ramkumar J, Nguyen DB;
PI Tribouley CM, Lo TP, Au-Young J, Thangavelu K, Kearney L;
XX WPI: 2002-463471/49.
DR N-PSDB; ABK86140.
XX
XX
PT New human proteases useful for diagnosing, preventing or treating
PT anorexia, myocardial infarction, Addison's disease, hepatitis, Cushing's
PT syndrome, eczema, Parkinson's disease, and impotence.
XX
XX
PS Claim 1; Page 140-143; 168pp; English.
XX
XX This invention relates to the DNA and protein sequences of novel isolated
XX human proteases (PRTS). the protein sequences of the invention are useful
CC for screening a compound for effectiveness as an agonist or antagonist of
CC its activity, the identified agonist and antagonist are useful for
CC treating a disease or condition associated with decreased or
CC overexpression of functional PRTS in a patient. The PRTS protein is also
CC useful as an immunogen for preparing polyclonal or monoclonal antibodies
CC by hybridoma technology. An antibody that binds the PRTS proteins is
CC useful for detection and purification of the proteins and can be used to
CC diagnose a condition or disease associated with expression of PRTS in a
CC subject or in a biological sample. The sequences of the invention are
CC useful for diagnosis, treatment and prevention of gastrointestinal
CC disorders such as gastritis, ulcerative colitis, Reye's syndrome, etc;
CC cardiovascular such as atherosclerosis, hypertension, myocardial
CC infarction, etc; autoimmune/inflammatory disorders such as acquired
CC immunodeficiency syndrome (AIDS), asthma, grave's disease, etc; cell
CC proliferative disorders such as hepatitis, psoriasis, leukemia etc;
CC developmental disorders such as Cushing's syndrome; epithelial disorders
CC such as dermatitis, scabies, eczema, etc; neurological disorders such as
CC Parkinson's disease, dementia, Alzheimer's disease, Huntington's disease,
CC multiple sclerosis, etc; or reproductive disorders such as infertility,
CC impotence. The present sequence represents the PRTS7 protein sequence of
XX the invention
XX
SQ Sequence 1251 AA;

Query Match 85.9%; Score 666; DB 5; Length 1251;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 766; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTIVDKASESSDPSAYQNQGSSEAVSPGMDAGSASGAVSSLNDVSNHTLSLGPVGA 60
DB 1 MTIVDKASESSDPSAYQNQGSSEAVSPGMDAGSASGAVSSLNDVSNHTLSLGPVGA 60

QY 61 VVYSSSVDPKSPQKQALGDGIAPPQKVLFPSEKICLKQOOTHVRVAGLQNLGNTC 120
DB 61 VVYSSSVDPKSPQKQALGDGIAPPQKVLFPSEKICLKQOOTHVRVAGLQNLGNTC 120

QY 121 FANAALQCLTYTTPLANYMLSHEHSKTCCHAEFCMCTMQAHI7QALSNPGDVIKPMFVI 180
DB 121 FANAALQCLTYTTPLANYMLSHEHSKTCCHAEFCMCTMQAHI7QALSNPGDVIKPMFVI 180

QY 181 NEMRRIARHLRFQNGEDAHEFLQYTVTDAMQKACLNGSNKLDHRHTQATTLVCQIFGGYILRS 240
DB 181 NEMRRIARHLRFQNGEDAHEFLQYTVTDAMQKACLNGSNKLDHRHTQATTLVCQIFGGYILRS 240

QY 241 RVKCLNCKGVSDTFDPLYDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKCSCKKWVPA 300
DB 241 RVKCLNCKGVSDTFDPLYDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKCSCKKWVPA 300

QY 301 SKRFTTHRSNSVLTLSLKRFAFTGGKIADVKYPEYLDIRPYMSQNGEPIVYVLAVL 360
DB 301 SKRFTTHRSNSVLTLSLKRFAFTGGKIADVKYPEYLDIRPYMSQNGEPIVYVLAVL 360

QY 361 VHTGFNCHAGHYFCYIKASNGLVQMNDISIVTSDIRSVLSQQAIVLYFIRSHDVKNNGE 420
DB 361 VHTGFNCHAGHYFCYIKASNGLVQMNDISIVTSDIRSVLSQQAIVLYFIRSHDVKNNGE 420

QY 421 LTHPTSPQSSPRPVISQRVTVNKAQPGFQGLPFSHMIKRPPLHNGTGPLKDTSPSS 480

Db	421	LTHPTSPGSSPRVVISQRVVYTNQAAPGFLCPQJPSHMIKNPHLNGTGPLKOTPPSS	480
QY	481	MSSPNGNSVNRASPVNASASVQVNSVNRSSVPIPEHPKKQKITIISHNKLPIVRQCSQEN	540
Db	481	MSSPNGNSVNRASPVNASASVQVNSVNRSSVPIPEHPKKQKITIISHNKLPIVRQCSQEN	540
QY	541	LHSNSLENTKPVPSSTITNSAVOSTSNASTMSVSSKVTKPIPRSESCSQPYMNGSKSLN	600
Db	541	LHSNSLENTKPVPSSTITNSAVOSTSNASTMSVSSKVTKPIPRSESCSQPYMNGSKSLN	600
QY	601	SSVLVPYGAESSESDSEBSKGLGKENGIGTIVSSHSPGQDAEDEATPHEIQEPMTLNGA	660
Db	601	SSVLVPYGAESSESDSEBSKGLGKENGIGTIVSSHSPGQDAEDEATPHEIQEPMTLNGA	660
QY	661	NSAUSDSDPKENGILAPDGASCOGQPALHSENPFAKANGLPGLMPAPLLSLPEDKILETF	720
Db	661	NSAUSDSDPKENGILAPDGASCOGQPALHSENPFAKANGLPGLMPAPLLSLPEDKILETF	720
QY	721	RLSNKLKGSTDEMSPGAERGGPPEDRDAPQPGSPAESLEPPDAAA	767
Db	721	RLSNKLKGSTDEMSPGAERGGPPEDRDAPQPGSPAESLEPPDAAA	767
RESULT 3			
AAB82177			
ID	AAB82177 standard; protein; 762 AA.		
XX	AAB82177;		
AC	AAB82177;		
DT	20-JUL-2001 (first entry)		
XX	Human ubiquitin protease 23431.		
DE	Human; ubiquitin protease; deubiquitinating enzyme; tuberculostatic;		
XX	antiaesthetic; antiinflammatory; antidiarrhoeic; hepatotropic;		
KW	gynaecological; cytostatic; antimicrobial; neuroprotective; anti-HIV;		
KW	immunosuppressive; cardiact; antianaemic; nephrotropic; antibacterial;		
KW	anti-thyroid; gastrointestinal.		
XX	Homo sapiens.		
OS	Homo sapiens.		
XX	WO2001233589-A2.		
FN	WO2001233589-A2.		
PD	05-APR-2001.		
XX	29-SEP-2000; 2000WO-US026962.		
PF	29-SEP-1999; 99US-00407356.		
XX	(MILL-) MILLENNIUM PHARM INC.		
PA	Kapeller-Libermann R;		
XX	WPI; 2001-374253/39.		
PI	N-PSDB; AAH19322.		
XX	New human ubiquitin protease, a member of the mammalian deubiquitinating		
PT	enzymes is useful for diagnosis and treatment of e.g. tuberculosis and		
PT	Alzheimer's disease.		
XX	Claim 5; Fig 1; 111pp; English.		
PS	The present sequence is human ubiquitin protease 23431. The ubiquitin		
XX	protease coding sequence and protein are useful as reagents or targets in		
CC	assays for treatment and diagnosis of ubiquitin-mediated or -related		
CC	disorders, especially disorders mediated by deubiquitinating enzymes. The		
CC	protein and coding sequence are also useful for treating disorders		
CC	involving the following: the spleen e.g. tuberculosis and typhoid fever,		
CC	the lung such as bronchial asthma, the colon e.g. as diarrhoea and		
CC	dysentery, the liver e.g. jaundice and cholestasis, the uterus and		
CC	endometrium e.g. endometritis, the brain e.g. acute meningitis and		
CC	Alzheimer disease. T-cells including T-lymphocytes, T-helper cells and		

New human ubiquitin protease, a member of the mammalian deubiquitinating enzymes is useful for diagnosis and treatment of e.g. tuberculosis and Alzheimer's disease.

Claim 5; Fig 1; 111pp; English.

The present sequence is human ubiquitin protease 23431. The ubiquitin protease coding sequence and protein are useful as reagents or targets in assays for treatment and diagnosis of ubiquitin-mediated or -related disorders, especially disorders mediated by deubiquitinating enzymes. The protein and coding sequence are also useful for treating disorders involving the following: the spleen e.g. tuberculosis and typhoid fever, the lung such as bronchial asthma, the colon e.g. as diarrhoea and dysentery, the liver e.g. jaundice and cholestasis, the uterus and endometrium e.g. endometritis, the brain e.g. acute meningitis and Alzheimer disease, T-cells including transplant rejection and autoimmune

CC diseases such as systemic lupus erythematosus, diseases of the skin such
CC as malignant melanoma, the bone marrow e.g. B- and T-lymphoid leukaemias,
CC the heart including myocardial infarction, red cells e.g. anaemia, the
CC thymus e.g. Hodgkin disease, B-cells e.g. peripheral B-cell neoplasms,
CC the kidney e.g. polycystic kidney disease, the breast including
CC periductal mastitis, the testis and epididymis e.g. syphilis, the
CC prostate e.g. nodular hyperplasia, the thyroid, e.g. hyperthyroidism, the
CC skeletal muscle e.g. rhabdomyosarcoma, the pancreas e.g. ectopic
CC pancreas, reduced platelet number e.g. HIV (human immunodeficiency virus)
CC -associated thrombocytopaenia and disorders involving precursor T-cell
CC neoplasms including precursor T lymphoblastic leukaemia/lymphoma
XX
SQ Sequence 762 AA;
Query Match 85.3%; Score 661; DB 4; Length 762;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 761; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MTIVDKASSSDPFAYONQPSSEAVSPGDMAGSAGVSSINDVSNHTLSLGPVPGA 60
DB 1 MTIVDKASSSDPFAYONQPSSEAVSPGDMAGSAGVSSINDVSNHTLSLGPVPGA 60
QY 61 VVYSSSVDPKSPKQDQALGDGIAPPOKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
DB 61 VVYSSSVDPKSPKQDQALGDGIAPPOKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTTPPLANYMLSHEHSKTCBAGFCMCTQOAHITQALSNPDGVDVKMFVI 180
DB 121 FANAALQCLTYTTPPLANYMLSHEHSKTCBAGFCMCTQOAHITQALSNPDGVDVKMFVI 180
QY 181 NEMRRITARHFRFGNQEDAHEFLQYTDVAMQACINGSNKLDRTHTQATTLVCOIFGGVYLS 240
DB 181 NEMRRITARHFRFGNQEDAHEFLQYTDVAMQACINGSNKLDRTHTQATTLVCOIFGGVYLS 240
QY 241 RVKCLCKGVSDFDPVLDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKCSKCKMWPA 300
DB 241 RVKCLCKGVSDFDPVLDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKCSKCKMWPA 300
QY 301 SKRFTTHRSSNVLTLSLKRFANFTGGKIAKDVKYPEYLDIRPYMSQNGEPFVVLYAVL 360
DB 301 SKRFTTHRSSNVLTLSLKRFANFTGGKIAKDVKYPEYLDIRPYMSQNGEPFVVLYAVL 360
QY 361 VHTGFNCHAGHYCYIKASNGLVQWMDNSIVSTDSVLSQQAAYVLFYIRSHDVKNNGE 420
DB 361 VHTGFNCHAGHYCYIKASNGLVQWMDNSIVSTDSVLSQQAAYVLFYIRSHDVKNNGE 420
QY 421 LTHPTSPGSSPRPVTSQRVVTNKQAAPGFIGPQLPSHMIKNPPHNLGTGPKLDTFSSS 480
DB 421 LTHPTSPGSSPRPVTSQRVVTNKQAAPGFIGPQLPSHMIKNPPHNLGTGPKLDTFSSS 480
QY 481 MSSPNGSSVNRASPVNASQVNSVNRSSVIPHEPKKQKITITISHNKLPVRCQSQPN 540
DB 481 MSSPNGSSVNRASPVNASQVNSVNRSSVIPHEPKKQKITITISHNKLPVRCQSQPN 540
QY 541 LHSNLENPKVPSSITTSNAVOSTSNASTMSVSSVTKVPIPRSESCSQPVNMGSKLN 600
DB 541 LHSNLENPKVPSSITTSNAVOSTSNASTMSVSSVTKVPIPRSESCSQPVNMGSKLN 600
QY 601 SSVLPVYGASSEDSDSESKGLGKENGITGVTSVSHSPQDAEDBEATPHELPMTLNGA 660
DB 601 SSVLPVYGASSEDSDSESKGLGKENGITGVTSVSHSPQDAEDBEATPHELPMTLNGA 660
QY 661 NSADSDPKENGLAPDGASQCGQPALHSENPFAKNGLPKMLPAPLISPEDKILETF 720
DB 661 NSADSDPKENGLAPDGASQCGQPALHSENPFAKNGLPKMLPAPLISPEDKILETF 720
QY 721 RLSNKLKGTDMSPAGRGPEDRDAPQPGSPAESLEE 762
DB 721 RLSNKLKGTDMSPAGRGPEDRDAPQPGSPAESLEE 762

ID XX ABU08951 standard; prote.n; 762 AA.
AC ABU08951;
XX 10-JUN-2003 (first entry)
DT Human ubiquitin protease.
DE Human; enzyme; ubiquitin protease; spleen disorder; Hodgkin's disease;
KW lung disorder; adult respiratory distress syndrome; colon disorder;
KW inflammatory bowel disease; liver disorder; jaundice; uterine disorder;
KW endometriosis; brain disorder; Alzheimer's disease; T-cell disorder;
KW acquired immunodeficiency syndrome; AIDS; skin disorder; urticaria;
KW heart disorder; ischaemic heart disease; blood vessel disorder;
KW atherosclerosis; red blood cell disorder; anaemia; thymus disorder;
KW DiGeorge syndrome; B-cell disorder; leukaemia; kidney disorder;
KW polycystic kidney disease; glomerulonephritis; breast disorder; mastitis;
KW testicular disorder; sexually transmitted disease; thyroid disorder;
KW hypothyroidism; pancreatic disorder; pancreatitis; intestinal disorder;
KW whipple disease; tumour; cancer.
XX Homo sapiens.
OS US2003037350-A1.
XX 05-JUN-2002; 2002US-00163547.
XX 29-SEP-1999; 99US-00407356.
XX 05-NOV-1999; 99US-00435311.
PR 01-FEB-2000; 2000US-00496005.
PR 11-FEB-2000; 2000US-0182009P.
PR 14-FEB-2000; 2000US-0182408P.
PR 28-FEB-2000; 2000US-0185503P.
PR 02-NOV-2000; 2000US-00704918.
PR 12-FEB-2001; 2001US-00781598.
PR 14-FEB-2001; 2001US-00782952.
PR 28-FEB-2001; 2001US-00796100.
XX (MILL-) MILLENNIUM PHARM INC.
XX Gluckmann MA, Kapeller-Libermann R, Meyers RE, Rudolph-Owen LA;
WI: 2003-342047/32.
N-PSDB; ABX93880.
XX New nucleic acid or polypeptide, useful for preparing a composition for
diagnosing or treating diseases e.g., tumor.
XX Claim 9; Fig 1; 281pp; English.
PS The invention relates to a new isolated nucleic acid molecule encoding
XX one of 6 polypeptides (ubiquitin protease, lipase, Dynamain, short chain
CC dehydrogenase, ADAM-17 (a disintegrin and metalloprotease domain protein
CC with thrombospondin (TS) domains) and gamma butyrobetaine-hydroxylase
CC (gamma BH)). Also included are the polypeptide, host cells containing the
CC nucleic acids, an antibody that selectively binds to the polypeptide, a
CC method for producing the polypeptide, a method for detecting the presence
CC of the polypeptide or the nucleic acid in a sample, a method for
CC identifying a compound that binds to the polypeptide, a method for
CC modulating the activity of the polypeptide and a method for identifying a
CC compound that modulates the activity of the polypeptide. The nucleic acid
CC or polypeptide is useful for preparing a composition for diagnosing or
CC treating diseases e.g. spleen disorders (e.g. splenomegaly and Hodgkin's
CC disease), lung disorders (e.g. adult respiratory distress syndrome,
CC pulmonary oedema, chronic bronchitis and emphysema), colon disorders
CC (e.g. stenosis, colitis, inflammatory bowel disease and Crohn's disease),
CC liver disorders (e.g. jaundice, cirrhosis, hepatitis and alcoholic liver
CC disease), uterine and endometrial disorders (e.g. endometriosis and
CC menopausal changes), brain disorders (e.g. encephalitis, Alzheimer's
CC disease, Parkinson's disease, ataxia and multiple sclerosis), T-cell
CC disorders (e.g. acquired immunodeficiency syndrome, AIDS), skin disorders

CC (e.g. urticaria, dermatitis and lupus erythematosus), heart disorders
 CC (e.g. ischaemic heart disease, myocardial infarction and cardiomyopathy),
 CC blood vessel disorders (e.g. atherosclerosis, thrombophlebitis and
 CC Raynaud disease), red blood cell disorders (e.g. anaemia), thymus
 CC disorders (e.g. DiGeorge syndrome), B-cell disorders (e.g. leukaemia),
 CC kidney disorders (e.g. polycystic kidney disease and glomerulonephritis),
 CC breast disorders (e.g. mastitis), testicular disorders (e.g. sexually
 CC transmitted diseases and cryptorchidism), thyroid disorders (e.g.
 CC hypothyroidism), pancreatic disorders (e.g. pancreatitis), and intestinal
 CC disorders (e.g. whipple disease), as well as tumours and cancers of the
 CC above listed organs/cells. Many more diseases and disorders are listed in
 CC the specification. The present sequence represents human ubiquitin
 CC protease
 XX
 SQ Sequence 762 AA;

Query Match 85.3%; Score 661; DB 6; Length 762;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 761; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTIVDKASESDPSAYONQPGSEAVSPGDMAGSAGVSSINDVSNHTLSLGPVPGA 60
 DB 1 MTIVDKASESDPSAYONQPGSEAVSPGDMAGSAGVSSINDVSNHTLSLGPVPGA 60

QY 61 VYSSSSVDPKSPKQKQALGDGIAPKVLFPSEKICLKQOQTHRVGAGLQNLGNTC 120
 DB 61 VYSSSSVDPKSPKQKQALGDGIAPKVLFPSEKICLKQOQTHRVGAGLQNLGNTC 120

QY 121 FANAALQCLTYTPPLANYMLSHESKTCCHAEFCMCTQOAHITQALSNPQGVKPMFVI 180
 DB 121 FANAALQCLTYTPPLANYMLSHESKTCCHAEFCMCTQOAHITQALSNPQGVKPMFVI 180

QY 181 NEMRIARHLRFGNQDEDAHEFLQYTVDMQKACLSNKLDRHTQATTLVCCQIFGCVLS 240
 DB 181 NEMRIARHLRFGNQDEDAHEFLQYTVDMQKACLSNKLDRHTQATTLVCCQIFGCVLS 240

QY 241 RVKCLNCKGVSTFDPYLDITLKAQSVNKALEQFVKPEQLDGENSYKCKCKKWPA 300
 DB 241 RVKCLNCKGVSTFDPYLDITLKAQSVNKALEQFVKPEQLDGENSYKCKCKKWPA 300

QY 301 SKRFTIHRSSNVLTLSLKRFANFTGKIAKDYKPYLDIRPYMQPNGEPIVYVLYAVL 360
 DB 301 SKRFTIHRSSNVLTLSLKRFANFTGKIAKDYKPYLDIRPYMQPNGEPIVYVLYAVL 360

QY 361 VHTGPNCHAGHYFCYIKASNGELWYQWNSISVTSIRSVLSQAYVLFIRSHDVKNGE 420
 DB 361 VHTGPNCHAGHYFCYIKASNGELWYQWNSISVTSIRSVLSQAYVLFIRSHDVKNGE 420

QY 421 LTHPTSPGQSPRPVISCORVVVTKQAAPGFTGPQLPSHMIKNPPLHNGTGPKLQTPSS 480
 DB 421 LTHPTSPGQSPRPVISCORVVVTKQAAPGFTGPQLPSHMIKNPPLHNGTGPKLQTPSS 480

QY 481 MSSPNGNSVNRASPVNASVQWNSVNRSSVIFPHPKQKITISIHKNLPRVQCQSPN 540
 DB 481 MSSPNGNSVNRASPVNASVQWNSVNRSSVIFPHPKQKITISIHKNLPRVQCQSPN 540

QY 541 LHSNSLENPTKVPSSITTSNAVOSTSNASTWSVSKVTPIPRSESCSQVPMGKSKLN 600
 DB 541 LHSNSLENPTKVPSSITTSNAVOSTSNASTWSVSKVTPIPRSESCSQVPMGKSKLN 600

QY 601 SSVLVAPGASESDDESGLKGENGIGTIVSSHSPQDAEDEATPHELQEPMTLNGA 660
 DB 601 SSVLVAPGASESDDESGLKGENGIGTIVSSHSPQDAEDEATPHELQEPMTLNGA 660

QY 661 NSADSDSDPKENGLAPDGCOCOPALHSENTPFAKANGLPKMLPAPLLSPEDKILETF 720
 DB 661 NSADSDSDPKENGLAPDGCOCOPALHSENTPFAKANGLPKMLPAPLLSPEDKILETF 720

QY 721 RLNSNKLKGTDEMSAPGAERGFPPEDRDAEPQPGSPAAESLEE 762
 DB 721 RLNSNKLKGTDEMSAPGAERGFPPEDRDAEPQPGSPAAESLEE 762

RESULT 5
 AAB95146
 ID AAB95146 standard; protein; 1197 AA.
 XX
 AC AAB95146;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:17169.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 FN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-00116126.
 XX
 PR 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 03-JUN-2000; 2000JP-00241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 17169; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

Sequence 1197 AA;

Query Match 72.9%; Score 565; DB 4; Length 1197;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 765; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTIVDKASESDPSAYONQPGSEAVSPGDMAGSAGVSSINDVSNHTLSLGPVPGA 60
 DB 1 MTIVDKASESDPSAYONQPGSEAVSPGDMAGSAGVSSINDVSNHTLSLGPVPGA 60

QY 61 VYSSSSVPDKSKPSQKQDQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
DB 61 VYSSSSVPDKSKPSQKQDQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTPPLANYMLSHSESKTCHAEFCMMCTMQAHTQALSNPGDVVKPMFVI 180
DB 121 FANAALQCLTYTPPLANYMLSHSESKTCHAEFCMMCTMQAHTQALSNPGDVVKPMFVI 180
QY 181 NEMRIARHLRFGNQEDAHEFLQYTVDMQKACLSNGSKLDRHTQATTLVCOIFGGLYLS 240
DB 181 NEMRIARHLRFGNQEDAHEFLQYTVDMQKACLSNGSKLDRHTQATTLVCOIFGGLYLS 240
QY 241 RVKCLNCKGVSTDFPYLDITLLEIKAAQSVNKALEQFVKPEQLDGENSKYKCKCKKWPA 300
DB 241 RVKCLNCKGVSTDFPYLDITLLEIKAAQSVNKALEQFVKPEQLDGENSKYKCKCKKWPA 300
QY 301 SKRFTTHRSSNVLTLSLKRFPANFTGKIAKVPEYLDIRPMSOPNGEPIVYVLYAVL 360
DB 301 SKRFTTHRSSNVLTLSLKRFPANFTGKIAKVPEYLDIRPMSOPNGEPIVYVLYAVL 360
QY 361 VHTGFNCHAGHYFCYIKASNGLWYQMNDSIVSTSDIRSVLSQQAAYLVYIRSHDVKNNGE 420
DB 361 VHTGFNCHAGHYFCYIKASNGLWYQMNDSIVSTSDIRSVLSQQAAYLVYIRSHDVKNNGE 420
QY 421 LTHPTSPGQSSPRPVISQIRVWTKQAAPGFIGPQLPSHMIKNPPHLNGTGLKDTSPSS 480
DB 421 LTHPTSPGQSSPRPVISQIRVWTKQAAPGFIGPQLPSHMIKNPPHLNGTGLKDTSPSS 480
QY 481 MSSPFGNSSVNRASPVNASQVNSVNRSSVPEHPKOKITISHNKLPVRCQSQPN 540
DB 481 MSSPFGNSSVNRASPVNASQVNSVNRSSVPEHPKOKITISHNKLPVRCQSQPN 540
QY 541 LHSNLENPKVPSSITNSAVOSTSNASTMSVSSKVTPIPRSESCSPVWNGSKLIN 600
DB 541 LHSNLENPKVPSSITNSAVOSTSNASTMSVSSKVTPIPRSESCSPVWNGSKLIN 600
QY 601 SSVLPVYGAESSEDSDESKGLKENGIGTIVSSHSPQDAEDENPHELOEPTLNGA 660
DB 601 SSVLPVYGAESSEDSDESKGLKENGIGTIVSSHSPQDAEDENPHELOEPTLNGA 660
QY 661 NSADSDPKENGLAPDGASCOGAPALHSENPFKAKNGLPKMLPAPLLSLPDKILETF 720
DB 661 NSADSDPKENGLAPDGASCOGAPALHSENPFKAKNGLPKMLPAPLLSLPDKILETF 720
QY 721 RLSNKLKGSTDEMSAPGAERPPEDRDAPQPGSPAESLEPDAAA 767
DB 721 RLSNKLKGSTDEMSAPGAERPPEDRDAPQPGSPAESLEPDAAA 767

RESULT 6
ABP69291
ID ABP69291 standard; protein; 1270 AA.
XX AC ABP69291;
XX DE Human polypeptide SEQ ID NO 1338.
XX 20-JAN-2003 (first entry)
XX Human; genome mapping; gene therapy; food supplement; virus; fungus;
XX cell-proliferative disorder; neurodegenerative disease; bacterial;
XX Parkinson's disease; Alzheimer's disease; autoimmune disease;
XX multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
XX arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
XX antiParkinsonian; antidiabetic; immunosuppressive; dermatological;
XX haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
XX antiarthritic.
XX Homo sapiens.
XX WO200270539-A2.
XX 12-SEP-2002.
PD

XX 05-MAR-2002; 2002WO-US0005:095.
XX 05-MAR-2001; 2001US-00799451.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
XX Wehrman T, Wang J, Wang D, Drmanac RT;
XX WPI; 2002-759812/82.
XX N-PSDB; ABZ11508.
XX New polynucleotides comprising sequences assembled from expressed
XX sequence tags (ESTs), useful for treating cell-proliferative,
XX neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
XX or coagulation disorders.
XX Claim 9; SEQ ID NO 1338; 1012pp + Sequence Listing; English.
XX The invention relates to an isolated polynucleotide (I) comprising a
XX nucleotide sequence selected from any of 948 sequences (ABZ11119-
XX ABZ12066) or their mature protein coding portion, active domain coding
XX protein or complementary sequences. The polynucleotides are useful for
XX identifying expressed genes or for physical mapping of human genome. The
XX encoded polypeptides (ABP6892-ABP6949) are useful as molecular weight
XX markers, as a food supplement, for generating antibodies, in medical
XX imaging, screening and diagnostic assays and for treating cell-
XX proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
XX or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
XX diabetes, lupus) genetic disorders, myeloid or lymphoid disorders, liver
XX or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
XX or arthritis, etc. Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 1270 AA;

Query Match 72 9%; Score 565; DB 5; Length 1270;
Best Local Similarity 99 7%; Pred. No. 0;
Matches 765; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MTIVDKASESDPSAYQNPQGSSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVGA 60
DB 1 MTIVDKASESDPSAYQNPQGSSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVGA 60
QY 61 VYSSSSVPDKSKPSQKQDQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
DB 61 VYSSSSVPDKSKPSQKQDQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTPPLANYMLSHSESKTCHAEFCMMCTMQAHTQALSNPGDVVKPMFVI 180
DB 121 FANAALQCLTYTPPLANYMLSHSESKTCHAEFCMMCTMQAHTQALSNPGDVVKPMFVI 180
QY 181 NEMRIARHLRFGNQEDAHEFLQYTVDMQKACLSNGSKLDRHTQATTLVCOIFGGLYLS 240
DB 181 NEMRIARHLRFGNQEDAHEFLQYTVDMQKACLSNGSKLDRHTQATTLVCOIFGGLYLS 240
QY 241 RVKCLNCKGVSTDFPYLDITLLEIKAAQSVNKALEQFVKPEQLDGENSKYKCKCKKWPA 300
DB 241 RVKCLNCKGVSTDFPYLDITLLEIKAAQSVNKALEQFVKPEQLDGENSKYKCKCKKWPA 300
QY 301 SKRFTTHRSSNVLTLSLKRFPANFTGKIAKVPEYLDIRPMSOPNGEPIVYVLYAVL 360
DB 301 SKRFTTHRSSNVLTLSLKRFPANFTGKIAKVPEYLDIRPMSOPNGEPIVYVLYAVL 360
QY 361 VHTGFNCHAGHYFCYIKASNGLWYQMNDSIVSTSDIRSVLSQQAAYLVYIRSHDVKNNGE 420
DB 361 VHTGFNCHAGHYFCYIKASNGLWYQMNDSIVSTSDIRSVLSQQAAYLVYIRSHDVKNNGE 420
QY 421 LTHPTSPGQSSPRPVISQIRVWTKQAAPGFIGPQLPSHMIKNPPHLNGTGLKDTSPSS 480

Db 421 LTHPTSPGSSPRPVISQRVVTNKQAAPGFIGQLPSHMIKNPPLHNGTGPKLDTSSS 480
 QY MSSPNGNSVNRASPVNASVQVNSVNRSGVIEHPKPKQKITIISHNKLPVRCQSQPN 540
 Db 481 MSSPNGNSVNRASPVNASVQVNSVNRSGVIEHPKPKQKITIISHNKLPVRCQSQPN 540
 QY LHSNLSNLTPTKPVPSSTITNSAVQSTNSASTMSVSSKVTPIPRSESCSPVNMGSKLN 600
 Db 541 LHSNLSNLTPTKPVPSSTITNSAVQSTNSASTMSVSSKVTPIPRSESCSPVNMGSKLN 600
 QY SSVLVYPGAESESDSEESKGLKENGIGTIVSSHSPQDAEDEATPHELQEPMTLNGA 660
 Db 601 SSVLVYPGAESESDSEESKGLKENGIGTIVSSHSPQDAEDEATPHELQEPMTLNGA 660
 QY NSADSDSDPKENGLAPDGAQCQOPALHSENPFKANGLPGLKMPAPLLSLPDKILETF 720
 Db 661 NSADSDSDPKENGLAPDGAQCQOPALHSENPFKANGLPGLKMPAPLLSLPDKILETF 720
 QY 721 RLSNKLKSTDEMSAFGAERGPPEDRDAEPQGPSAAESLEEDAAA 767
 Db 721 RLSNKLKSTDEMSAFGAERGPPEDRDAEPQGPSAAESLEEDAAA 767

RESULT 7

ID ABB06118
 AB06118 standard; protein; 558 AA.

AC ABB06118;

DT 10-MAY-2002 (first entry)

XX Human NS protein sequence SEQ ID NO:210.

DE Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX antirheumatic; antiarthritic; antipsoriatic; ophthalmological; anti-HIV;

KW vasotrophic; antiarteriosclerotic; antiinflammatory; dermatological;

KW anorectic; muscular; antifertility; cardiovascular; anticoagulant;

KW antifibrinolytic; hypotension; antiasthmatic; immunomodulator; cardiac;

KW anticonvulsant; antidiabetic; tranquiliser; antidepressant; auroleptic;

KW gastrointestinal; virucide; antiulcer; cerebroprotective; nootropic;

KW contraceptive; vaccine; gene therapy; cancer; osteoporosis; dystonia;

KW endometrioid arthritis; degenerative disease; multiple sclerosis; psoriasis;

KW rheumatoid arthritis; cataract; restenosis; atherosclerosis; glaucoma;

KW inflammation; skin disorder; obesity; muscular dystrophy; AIDS;

KW infertility; cardiovascular disease; coagulation disease; hypertension;

KW ischaemia; asthma; immune disease; epilepsy; angina; neurodegeneration;

KW diabetes; anxiety; depression; schizophrenia; viral disease; stroke;

KW gastric ulcer; Alzheimer's disease.

XX Homo sapiens.

XX WO200206315-A2.

XX 24-JAN-2002.

XX 17-JUL-2001; 2001WO-IL000653.

XX 18-JUL-2000; 2000IL-00137345.

XX 15-DEC-2000; 2000IL-00140354.

XX (COMP-) COMPUGEN LTD.

XX Mintz L, Freilich S, Bernstein J;

XX WPI; 2002-155037/20.

XX N-PSDB; ABL39772.

XX One hundred and twenty eight novel nucleic acid sequences, useful for

XX treating and diagnosing e.g. cancer, asthma and Alzheimer's.

XX Claim 6; Page 243-245; 290pp; English.

XX

CC ABL39691 to ABL39818 represent novel human nucleic acid sequences
 CC encoding the proteins given in ABB06037 to ABB06164. The novel sequences
 CC (NS) can have cytostatic, osteopathic, gynaecological, neuroprotective,
 CC antirheumatic, antiarthritic, antipsoriatic, ophthalmological, virucide,
 CC vasotrophic, antiarteriosclerotic, antiinflammatory, dermatological,
 CC anorectic, muscular, anti-HIV, antifertility, cardiovascular,
 CC anticoagulant, antifibrinolytic, hypotension, antiasthmatic, cardiac,
 CC immunomodulator, anticonvulsant, antidiabetic, tranquiliser, antiulcer,
 CC antidepressant, gastrointestinal, auroleptic, cerebroprotective,
 CC nootropic and contraceptive activities. The NS can be used in vaccines,
 CC gene therapy and antitense therapy. Nucleic acids, expression vectors and
 CC antibodies from the present invention can be used for treating and
 CC diagnosing e.g. cancer, osteoporosis, endometriosis, degenerative
 CC diseases, dystonia, multiple sclerosis, rheumatoid arthritis, psoriasis,
 CC cataracts, restenosis, atherosclerosis, inflammation, skin disorders,
 CC glaucoma, obesity, muscular dystrophy, AIDS, infertility, cardiovascular
 CC disease, coagulation disease, ischaemia, hypertension, asthma, immune
 CC disease, epilepsy, angina, neurodegeneration, diabetes, anxiety,
 CC depression, schizophrenia, viral disease, gastric ulcers, stroke,
 CC Alzheimer's disease and as a contraceptive
 XX Sequence 558 AA;

Query Match

Best Local Similarity 57.8%; Score 448; DB 5; Length 558;

Matches 448; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 319 RFANFTGGKIAKDVKYPEYLDIRPYMSQNGEPIVYLVAVLVHTGFCNCHAGHYCYIKA 378

Db 111 RFANFTGGKIAKDVKYPEYLDIRPYMSQNGEPIVYLVAVLVHTGFCNCHAGHYCYIKA 170

QY 379 SNGLWYQMNDSIVSTDIRSVLSQAYVLYIRSHDVKNKGELTHTPSGQSSPRPVIS 438

Db 171 SNGLWYQMNDSIVSTDIRSVLSQAYVLYIRSHDVKNKGELTHTPSGQSSPRPVIS 230

QY 439 QRVTNKAQAPGFIGQLPSHMIKNPPLHNGTGPKLDTSSSMSSPNGNSVNRASPVNA 498

Db 231 QRVTNKAQAPGFIGQLPSHMIKNPPLHNGTGPKLDTSSSMSSPNGNSVNRASPVNA 290

QY 499 SASVQVNSVNRSSVIEHPKPKQKITIISHNKLPVRCQSQOPNLHNSLENTPKVPSSTI 558

Db 291 SASVQVNSVNRSSVIEHPKPKQKITIISHNKLPVRCQSQOPNLHNSLENTPKVPSSTI 350

QY 559 TNSAVQSTNSASTMSVSSKVTPIPRSESCSPVNMGSKLNSSVLVYPGAESESDSE 618

Db 351 TNSAVQSTNSASTMSVSSKVTPIPRSESCSPVNMGSKLNSSVLVYPGAESESDSE 410

QY 619 SKGLKENGIGTIVSSHSPQDAEDEATPHELQEPMTLNGANSADSDSDPKENGLAPDG 678

Db 411 SKGLKENGIGTIVSSHSPQDAEDEATPHELQEPMTLNGANSADSDSDPKENGLAPDG 470

QY 679 ASCQGPALHSENFPAKANGLPGLKMPAPLLSLPDKILETFRLSNKLKSTDEMSAPGA 738

Db 471 ASCQGPALHSENFPAKANGLPGLKMPAPLLSLPDKILETFRLSNKLKSTDEMSAPGA 530

QY 739 ERGPPEDRDAEPQGPSAAESLEEDAAA 766

Db 531 ERGPPEDRDAEPQGPSAAESLEEDAAA 558

RESULT 8

AAU23207

ID AAU23207 standard; protein; 156 AA.

XX AAU23207;

XX AAU23207;

DT 17-DEC-2001 (first entry)

XX Novel human enzyme polypeptide #293.

DE Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;

XX ligase; hyperproliferative disorder; immunodeficiency disorder;

KW autoimmune disorder; neurological disorder; metabolic disorder;

KW

Wed Aug 18 13:53:03 2004

us-10-049-745-4.oli10.rag

KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
 KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
 KW nephrotropic; anticoagulant.

OS Homo sapiens.

PN WO20015301-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001239.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 11-JUL-2000; 2000US-0218290P.

XX 14-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

XX 26-JUL-2000; 2000US-0224518P.

XX 14-AUG-2000; 2000US-0224519P.

XX 14-AUG-2000; 2000US-0225213P.

XX 14-AUG-2000; 2000US-0225214P.

XX 14-AUG-2000; 2000US-0225266P.

XX 14-AUG-2000; 2000US-0225267P.

XX 14-AUG-2000; 2000US-0225268P.

XX 14-AUG-2000; 2000US-0225270P.

XX 14-AUG-2000; 2000US-0225447P.

XX 14-AUG-2000; 2000US-0225757P.

XX 14-AUG-2000; 2000US-0225758P.

XX 14-AUG-2000; 2000US-0225759P.

XX 18-AUG-2000; 2000US-0226279P.

XX 22-AUG-2000; 2000US-0226681P.

XX 22-AUG-2000; 2000US-0226868P.

XX 22-AUG-2000; 2000US-0227182P.

PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236127P.
 PR 29-SEP-2000; 2000US-0236167P.
 PR 29-SEP-2000; 2000US-0236168P.
 PR 29-SEP-2000; 2000US-0236169P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236302P.
 PR 02-OCT-2000; 2000US-0237337P.
 PR 02-OCT-2000; 2000US-0237338P.
 PR 02-OCT-2000; 2000US-0237339P.
 PR 13-OCT-2000; 2000US-0239335P.
 PR 13-OCT-2000; 2000US-0239337P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0241826P.
 PR 08-NOV-2000; 2000US-024617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250191P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.


```
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR (HUMA-) HUMAN GENOME SCI INC.
XX
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX
XX DR WFI; 2001-465566/50.
XX DR N-PSDB; AAS41077.
XX
XX Novel polypeptides and polynucleotides useful for diagnosing, preventing,
PT treating neural, immune system, muscular, reproductive, pulmonary,
PT cardiovascular, renal, proliferative disorders and cancerous diseases.
XX
XX Claim 11; SEQ ID NO 1203; 1180pp; English.
XX
XX The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.
CC arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic
CC disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),
CC cardiovascular disorders (e.g. atherosclerosis), blood-related disorders
CC (e.g. haemophilia), reproductive disorders (e.g. infertility) and
CC infectious disorders (e.g. influenza). The polynucleotides of the
CC invention can also be used in gene therapy. AAU22915-AAU23814 represent
CC the novel human enzyme polypeptides of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 156 AA;
SQ
Query Match 19.1%; Score 148; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 6.7e-141; Indels 0; Gaps 0;
Matches 148; Conservative 0; Mismatches 0;
QY 1 MTIVDKAESDPSAYQNPQGSSEAVSPGDMGASGAVSLNDVSNHTLSLGFVPGA 60
Db 1 MTIVDKAESDPSAYQNPQGSSEAVSPGDMGASGAVSLNDVSNHTLSLGFVPGA 61
QY 61 VVYSSSVDPKSKPSQKQDQALGDGIAPPOKVLFPSEKICLKWOOTHRVGAGLQNLGNTC 120
Db 62 VVYSSSVDPKSKPSQKQDQALGDGIAPPOKVLFPSEKICLKWOOTHRVGAGLQNLGNTC 121
QY 121 FANAALQCLITYTPPIANMLSHEHSKTC 148
Db 122 FANAALQCLITYTPPIANMLSHEHSKTC 149
RESULT 9
AAU23743
ID AAU23743 standard; protein; 157 AA.
XX
XX AAU23743;
XX
XX 18-DEC-2001 (first entry)
XX
XX Novel human enzyme polypeptide #829.
XX
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
XX Homo sapiens.
OS
XX
```

WO200155301-A2.

02-AUG-2001.

17-JAN-2001; 2001WO-US001239.

31-JAN-2000; 2000US-0179065P.

04-FEB-2000; 2000US-0180628P.

24-FEB-2000; 2000US-0184664P.

02-MAR-2000; 2000US-0186350P.

16-MAR-2000; 2000US-0189874P.

17-MAR-2000; 2000US-0190076P.

18-APR-2000; 2000US-0198123P.

19-MAY-2000; 2000US-0205515P.

07-JUN-2000; 2000US-0209467P.

28-JUN-2000; 2000US-0214886P.

30-JUN-2000; 2000US-0215135P.

07-JUL-2000; 2000US-0216647P.

07-JUL-2000; 2000US-0216880P.

11-JUL-2000; 2000US-0217487P.

11-JUL-2000; 2000US-0217496P.

14-JUL-2000; 2000US-0218290P.

26-JUL-2000; 2000US-0220963P.

26-JUL-2000; 2000US-0220964P.

14-AUG-2000; 2000US-0224518P.

14-AUG-2000; 2000US-0225213P.

14-AUG-2000; 2000US-0225214P.

14-AUG-2000; 2000US-0225266P.

14-AUG-2000; 2000US-0225267P.

14-AUG-2000; 2000US-0225268P.

14-AUG-2000; 2000US-0225270P.

14-AUG-2000; 2000US-0225447P.

14-AUG-2000; 2000US-0225757P.

14-AUG-2000; 2000US-0225758P.

14-AUG-2000; 2000US-0225759P.

22-AUG-2000; 2000US-0226279P.

22-AUG-2000; 2000US-0226681P.

22-AUG-2000; 2000US-0226868P.

22-AUG-2000; 2000US-0227182P.

23-AUG-2000; 2000US-0227009P.

30-AUG-2000; 2000US-0228924P.

01-SEP-2000; 2000US-0229287P.

01-SEP-2000; 2000US-0229343P.

01-SEP-2000; 2000US-0229344P.

01-SEP-2000; 2000US-0229345P.

05-SEP-2000; 2000US-0229509P.

05-SEP-2000; 2000US-0229513P.

06-SEP-2000; 2000US-0230437P.

06-SEP-2000; 2000US-0230438P.

08-SEP-2000; 2000US-0231242P.

08-SEP-2000; 2000US-0231243P.

08-SEP-2000; 2000US-0231244P.

08-SEP-2000; 2000US-0231413P.

08-SEP-2000; 2000US-0231414P.

08-SEP-2000; 2000US-0232080P.

08-SEP-2000; 2000US-0232081P.

12-SEP-2000; 2000US-0231968P.

14-SEP-2000; 2000US-0232397P.

14-SEP-2000; 2000US-0232398P.

14-SEP-2000; 2000US-0232399P.

14-SEP-2000; 2000US-0232400P.

14-SEP-2000; 2000US-0232401P.

14-SEP-2000; 2000US-0233063P.

14-SEP-2000; 2000US-0233064P.

21-SEP-2000; 2000US-0233065P.

21-SEP-2000; 2000US-0234223P.

25-SEP-2000; 2000US-0234274P.

25-SEP-2000; 2000US-0234997P.

26-SEP-2000; 2000US-0234998P.

27-SEP-2000; 2000US-0235484P.

27-SEP-2000; 2000US-0235834P.

27-SEP-2000; 2000US-0235836P.

XX WPI: 2001-465566/50.
DR N-PSDB; AAS41613.
XX Novel polypeptides and polynucleotides useful for diagnosing, preventing,
PT treating neural, immune system, muscular, reproductive, pulmonary,
PT cardiovascular, renal, proliferative disorders and cancerous diseases.
XX Claim 11; SEQ ID NO 1739, 1180pp; English.
XX The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.
CC arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic
CC disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),
CC cardiovascular disorders (e.g. atherosclerosis), blood-related disorders
CC (e.g. haemophilia), reproductive disorders (e.g. infertility) and
CC infectious disorders (e.g. influenza). The polynucleotides of the
CC invention can also be used in gene therapy. AAU2915-AAU23814 represent
CC the novel human enzyme polypeptides of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 157 AA;
SQ
Query Match 16.5%; Score 128; DB 4; Length 157;
Best Local Similarity 100.0%; Pred. No. 1.2e-120;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MTIVDKASESSDPSAYQNPQGSSEAVSPGMDAGSASWGAIVSSINDVSNHTLSIGPVPGA 60
Db 2 MTIVDKASESSDPSAYQNPQGSSEAVSPGMDAGSASWGAIVSSINDVSNHTLSIGPVPGA 61
Qy 61 VVYSSSSVPDKSKPEPOKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
Db 62 VVYSSSSVPDKSKPEPOKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 121
Qy 121 FANAALQC 128
Db 122 FANAALQC 129
RESULT 10
AAO12240
ID AAO12240 standard; protein; 125 AA.
XX AAO12240;
XX AC AAO12240;
XX 06-NOV-2001 (first entry)
XX Human polypeptide SEQ ID NO 26132.
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX Homo sapiens.
OS WO200164835-A2.
XX PN
XX 07-SEP-2001.
XX PD
XX 26-FEB-2001; 2001WO-US004927.
XX PF
XX 28-FEB-2000; 2000US-00515126.
XX PR 18-MAY-2000; 2000US-00577409.
PR

PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX PA

```
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX N-PSDB; AAI92171.
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX and treating e.g. leukemia, inflammation and immune disorders.
XX Claim 20; SEQ ID NO 26132; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation. Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 125 AA;
XX
XX Query Match 16.0%; Score 124; DB 4; Length 125;
XX Best Local Similarity 100.0%; Pred. No. 1.1e-116;
XX Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MTIVKASESSDPSAYQNPQGSSEAVSPGMDAGASWGAVALNDVSNHTLSLGPVPGA 60
XX Db |||||
XX 2 MTIVKASESSDPSAYQNPQGSSEAVSPGMDAGASWGAVALNDVSNHTLSLGPVPGA 61
XX
XX QY 61 VVYSSSVDPKSPQKQDQALGDGIAPPQKVLFPSEKICLKWOOTHVRVAGLQNLGNTC 120
XX Db |||||
XX 62 VVYSSSVDPKSPQKQDQALGDGIAPPQKVLFPSEKICLKWOOTHVRVAGLQNLGNTC 121
XX
XX QY 121 FANA 124
XX Db |||||
XX 122 FANA 125
XX
XX RESULT 11
XX AAM16544
XX ID AAM16544 standard; protein; 112 AA.
XX AC AAM16544;
XX
XX DT 12-OCT-2001 (first entry)
XX
XX DE Peptide #2978 encoded by probe for measuring cervical gene expression.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer.
XX OS Homo sapiens.
XX
XX FN WO200157278-A2.
XX
XX Q9-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000670.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
```

```
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human cervical epithelial cells.
XX Claim 27; SEQ ID NO 21370; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs: see AAI10068-AAI28459). The present sequence is a peptide encoded
XX by one such probe. The SNPs are derived from human HeLa cells. The SENPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 112 AA;
XX
XX Query Match 14.5%; Score 112; DB 4; Length 112;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-104;
XX Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 591 PYMNGKSLNLSVLPVGYAESSEDSDEESKGLGKENGIGTIVSSHSPGQDADEEATPHE 650
XX Db |||||
XX 1 PYMNGKSLNLSVLPVGYAESSEDSDEESKGLGKENGIGTIVSSHSPGQDADEEATPHE 60
XX
XX QY 651 LOEPMTLGANSADSDPKENGLAPDGASCOCQOPALHSENFPAKANGLPKG 702
XX Db |||||
XX 61 LOEPMTLGANSADSDPKENGLAPDGASCOCQOPALHSENFPAKANGLPKG 112
XX
XX RESULT 12
XX ABB35528
XX ID ABB35528 standard; peptide; 112 AA.
XX XX ABB35528;
XX AC ABB35528;
XX
XX DT 04-FEB-2002 (first entry)
XX
XX DE Peptide #3034 encoded by human foetal liver single exon probe.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX OS Homo sapiens.
XX
XX FN WO200157277-A2.
XX
XX Q9-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000669.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
```

PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human fetal liver.

XX Claim 27; SEQ ID NO 28163; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human fetal liver. The
CC present sequence is a peptide encoded by a single exon nucleic acid probe
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 112 AA;

Query Match 14.5%; Score 112; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.4e-104;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 591 PVNMGSKLNSSVLPVYGAESESDSEESKGLGKENGIGTIVSSHSPGQDADEEATPHE 650
Db 1 PVNMGSKLNSSVLPVYGAESESDSEESKGLGKENGIGTIVSSHSPGQDADEEATPHE 60
QY 651 LQBPMTLNGANSADSDPKENGLAPDGASCOGQOPALHSENPFKANGLPK 702
Db 61 LQBPMTLNGANSADSDPKENGLAPDGASCOGQOPALHSENPFKANGLPK 112

RESULT 13

ABB30354
ID ABB30354 standard; peptide; 112 AA.

AC ABB30354;

XX 01-FEB-2002 (first entry)

XX Peptide #3005 encoded by breast cell single exon nucleic acid probe.

XX Human; microarray; single exon probe; gene expression; breast; disease;
XX cancer.

XX Homo sapiens.

XX WO200157271-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US0000662.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes, useful
XX for measuring gene expression in sample derived from human breast,
XX comprises number of single exon nucleic acid probes.

XX Claim 27; SEQ ID NO 13322; 327pp + Sequence Listing; English.

XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and BT 474 cells. The method involves contacting the

CC probes with a collection of detectably labelled nucleic acids derived
CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a peptide
CC encoded by a single exon nucleic acid probe of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 112 AA;

Query Match 14.5%; Score 112; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.4e-104;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 591 PVNMGSKLNSSVLPVYGAESESDSEESKGLGKENGIGTIVSSHSPGQDADEEATPHE 650
Db 1 PVNMGSKLNSSVLPVYGAESESDSEESKGLGKENGIGTIVSSHSPGQDADEEATPHE 60
QY 651 LQBPMTLNGANSADSDPKENGLAPDGASCOGQOPALHSENPFKANGLPK 702
Db 61 LQBPMTLNGANSADSDPKENGLAPDGASCOGQOPALHSENPFKANGLPK 112

RESULT 14

AAM68717

ID AAM68717 standard; protein; 112 AA.

AC AAM68717;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed probe encoded protein SEQ ID NO: 29023.

XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US0000668.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.

XX Example 4; SEQ ID NO 29023; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human

CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention
XX
SQ Sequence 112 AA;

Query Match 14.5%; Score 112; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.4e-104;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 591 PVWNGSKLNSVLPVPGAESSEDSDESKGLGKENGIGITIVSSHSPQDAEDDEATPHE 650
Db 1 PVWNGSKLNSVLPVPGAESSEDSDESKGLGKENGIGITIVSSHSPQDAEDDEATPHE 60
QY 651 LOEPMTLNGANSADSDPKENGLAPDAGSCQGPALHSENPFKANGLPCK 702
Db 61 LOEPMTLNGANSADSDPKENGLAPDAGSCQGPALHSENPFKANGLPCK 112

RESULT 15
AAM04259
ID AAM04259 standard; protein; 112 AA.
XX
AC AAM04259;
XX
DT 09-OCT-2001 (first entry)
XX
DE Peptide #2941 encoded by probe for measuring breast gene expression.
KW Probe; human; breast disease; breast cancer; development disorder;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US000661.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX
PR 26-MAY-2000; 2000US-0207456P.
XX
PR 30-JUN-2000; 2000US-00608408.
XX
PR 03-AUG-2000; 2000US-00632366.
XX
PR 21-SEP-2000; 2000US-0234687P.
XX
PR 27-SEP-2000; 2000US-0236359P.
XX
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
PT WPI; 2001-476286/51.
XX

Novel single exon nucleic acid probe used to measuring gene expression in
a human breast.
XX
Claim 27; SEQ ID NO 12999; 322pp; English.

The present invention relates to novel single exon nucleic acid probes
(see AAI00010-AA110067). The present sequence is a peptide encoded by one
such probe. The probes are useful for measuring human gene expression in
a human breast sample, where the probe hybridises at high stringency to a
nucleic acid expressed in the human breast. The probes are useful for
predicting, diagnosing, grading, staging, monitoring and prognosing
diseases of the human breast, particularly those diseases with polygenic
aetiology. The diseases include: breast cancer, disorders of polycystic
inflammatory diseases of the breast, fibrocystic changes, proliferative
breast disease and non-carcinoma tumours. Note: the sequence data for
this patent did not form part of the printed specification, but was
obtained in electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 112 AA;
Query Match 14.5%; Score 112; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.4e-104;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 591 PVWNGSKLNSVLPVPGAESSEDSDESKGLGKENGIGITIVSSHSPQDAEDDEATPHE 650
Db 1 PVWNGSKLNSVLPVPGAESSEDSDESKGLGKENGIGITIVSSHSPQDAEDDEATPHE 60
QY 651 LOEPMTLNGANSADSDPKENGLAPDAGSCQGPALHSENPFKANGLPCK 702
Db 61 LOEPMTLNGANSADSDPKENGLAPDAGSCQGPALHSENPFKANGLPCK 112

RESULT 16
ABG38302
ID ABG38302 standard; peptide; 112 AA.
XX
AC ABG38302;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human peptide encoded by genome-derived single exon probe SEQ ID 27967.
XX
KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US000665.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX
PR 26-MAY-2000; 2000US-0207456P.
XX
PR 30-JUN-2000; 2000US-00608408.
XX
PR 03-AUG-2000; 2000US-00632366.
XX
PR 21-SEP-2000; 2000US-0234687P.
XX
PR 27-SEP-2000; 2000US-0236359P.
XX
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
PT WPI; 2002-114183/15.
XX

Spatially-addressable set of single exon nucleic acid probes, used to
measure gene expression in human lung samples.
XX
Claim 27; SEQ ID NO 27967; 634pp; English.
XX
The invention relates to a spatially-addressable set of single exon
nucleic acid probes for measuring gene expression in a sample derived
from human lung comprising single exon nucleic acid probes having one of
12614 nucleic acid sequences mentioned in the specification, or their
complements or the 12387 open reading frames derived from the 12614
probes. Also included are a microarray comprising the novel set of probes
; the novel set of probes which hybridise at high stringency to a nucleic
acid expressed in the human lung; measuring gene expression in a sample
derived from human lung, comprising (a) contacting the array with a
collection of detectably labeled nucleic acids derived from human lung

XX	OS	Homo sapiens.	
XX	XX		
XX	PN	WO200157278-A2.	
XX	XX		
XX	PD	09-AUG-2001.	
XX	XX	<u>30-JAN-2001;</u>	2001WO-US000670.
XX	XX		
XX	PR	04-FEB-2000;	2000US-0180312P.
XX	PR	26-MAY-2000;	2000US-0207456P.
XX	ER	30-JUN-2000;	2000US-00608408.
XX	PR	03-AUG-2000;	2000US-00632366.
XX	PR	21-SEP-2000;	2000US-0234687P.
XX	PR	27-SEP-2000;	2000US-0236359P.
XX	PR	04-OCT-2000;	2000GB-00024263.
XX	XX		

XX	04-FEB-2000; 2000US-0180312P.
PR	26-MAY-2000; 2000US-0207458P.
PR	30-JUN-2000; 2000US-00608408.
PR	03-AUG-2000; 2000US-00632366.
PR	21-SEP-2000; 2000US-0234687P.
PR	27-SEP-2000; 2000US-0236359P.
PR	04-OCT-2000; 2000GB-00024263.
XX	
XX	(MOLE-) MOLECULAR DYNAMICS INC.
PA	
XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	
XX	WPI; 2001-483447/52.
XX	
XX	Human genome-derived single exon nucleic acid probes useful for analyzing
PT	gene expression in human fetal liver.
PT	

XX PS Claim 27; SEQ ID NO 30062; 639pp + Sequence Listing; English.

XX CC The invention relates to a single exon nucleic acid probe for measuring

CC human gene expression in a sample derived from human foetal liver. The

CC single exon nucleic acid probes may be used for predicting, measuring and

CC displaying gene expression in samples derived from human fetal liver. The

CC present sequence is a peptide encoded by a single exon nucleic acid probe

CC of the invention. Note: The sequence data for this patent did not form

CC part of the printed specification, but was obtained in electronic format

CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 66 AA;

Query Match 8.5%; Score 66; DB 4; Length 66;

Best Local Similarity 100.0%; Pred. No. 3.2e-58;

Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 LGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTTPPLANYMLS 141

Db 1 LGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTTPPLANYMLS 60

QY 142 HEHSKT 147

Db 61 HEHSKT 66

RESULT 19

AM30872

ID AAM30872 standard; protein; 66 AA.

AC AAM30872;

DT 17-OCT-2001 (first entry)

XX Peptide #4909 encoded by probe for measuring placental gene expression.

DE Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder.

KW Homo sapiens.

OS WO200157272-A2.

PN 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US0000663.

PF 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing

PT gene expression in human placenta.

XX Claim 27; SEQ ID NO 31141; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs:

CC see A131315-AA157546). The present sequence is a peptide encoded by one

CC such probe. The probes are useful for producing a microarray for

CC predicting, measuring and displaying gene expression in samples derived

CC from human placenta. The probes are useful for antenatal diagnosis of

CC human genetic disorders

XX SQ Sequence 66 AA;

Query Match 8.5%; Score 66; DB 4; Length 66;

Best Local Similarity 100.0%; Pred. No. 3.2e-58;

Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 LGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTTPPLANYMLS 141

Db 1 LGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTTPPLANYMLS 60

QY 142 HEHSKT 147

Db 61 HEHSKT 66

RESULT 20

ABB32185

ID ABB32185 standard; peptide; 66 AA.

XX ABB32185;

AC ABB32185;

DT 01-FEB-2002 (first entry)

XX Peptide #4836 encoded by breast cell single exon nucleic acid probe.

DE Human; microarray; single exon probe; gene expression; breast; disease;

XX cancer.

KW Homo sapiens.

OS WO200157271-A2.

PN 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US0000662.

PF 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI; 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes, useful

PT for measuring gene expression in sample derived from human breast,

PT comprises number of single exon nucleic acid probes.

XX Claim 27; SEQ ID NO 15153; 327pp + Sequence Listing; English.

XX The invention relates to a spatially-addressable set of single exon

CC nucleic acid probes for measuring gene expression in a sample derived

CC from human breast and BT 474 cells. The method involves contacting the

CC probes with a collection of detectably labelled nucleic acids derived

CC from mRNA of human breast, and then measuring the label bound to each

CC probe of the microarray. The probes are useful for verifying the

CC expression of regions of genomic DNA predicted to encode proteins. They

CC are useful for gene discovery, and for determining predisposition and/or

CC assessing the toxicity of chemical agents on cells. The microarray of

CC this invention presents a far greater diversity of probes for measuring

CC gene expression, with far less bias than expressed sequence tag

CC microarrays. The method is suitable for rapid production of functional

CC information from genomic sequence. The present sequence is a peptide

CC encoded by a single exon nucleic acid probe of the invention. Note: The

CC sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 66 AA;

Query Match
Best Local Similarity 100.0%; Score 66; DB 4; Length 66;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 LGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYMLS 141
DB 1 LGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYMLS 60
QY 142 HEHST 147
DB 61 HEHST 66

RESULT 21
ID AM70556
AC AM70556 standard; protein; 66 AA.
XX
XX AM70556;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 30862.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX Homo sapiens.
OS
XX WO200157276-A2.
PN
XX
PD 09-AUG-2001
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
PS Example 4; SEQ ID NO 30862; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention
XX
SQ Sequence 66 AA;

Query Match
Best Local Similarity 100.0%; Score 66; DB 4; Length 66;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 LGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYMLS 141
DB 1 LGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYMLS 60
QY 142 HEHST 147
DB 61 HEHST 66

DB 1 LGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYMLS 60
QY 142 HEHST 147
DB 61 HEHST 66

RESULT 22
ID ABG52236
AC ABG52236 standard; peptide; 66 AA.
XX
XX ABG52236;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver peptide, SEQ ID NO 30884.
XX
KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW hypercholesterolaemia; coronary heart disease.
XX
XX Homo sapiens.
OS
XX WO200157273-A2.
PN
XX
PD 09-AUG-2001
XX
XX 30-JAN-2001; 2001WO-US000664.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234587P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-48898/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
PS Claim 27; SEQ ID NO 30884; 658pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABG47348-ABG59930 represent human
CC liver single exon encoded peptides of the invention. Note: The sequence
CC information for this patent does not appear in the printed specification
CC but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 66 AA;

Query Match
Best Local Similarity 100.0%; Score 66; DB 4; Length 66;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 LGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYMLS 141
DB 1 LGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYMLS 60
QY 142 HEHST 147

Db 61 HEHST 66

RESULT 23
AAM05997
ID AAM05997 standard; protein; 66 AA.
XX
AC AAM05997;
XX
DT 09-OCT-2001 (first entry)
XX
DE Peptide #4679 encoded by probe for measuring breast gene expression.
KW Probe; human; breast disease; breast cancer; development disorder;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN WO200157270-A2.
PD
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US000661.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression in
XX a human breast.
XX
XX Claim 27; SEQ ID NO 14737; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes
XX (see AAL00010-AA110067). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for measuring human gene expression in
XX a human breast sample, where the probe hybridises at high stringency to a
XX nucleic acid expressed in the human breast. The probes are useful for
XX predicting, diagnosing, grading, staging, monitoring and prognosing
XX diseases of the human breast, particularly those diseases with polygenic
XX aetiology. The diseases include: breast cancer, disorders of development,
XX inflammatory diseases of the breast, fibrocystic changes, proliferative
XX breast disease and non-carcinoma tumours. Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 66 AA;
XX
XX Query Match 8.5%; Score 66; DB 4; Length 66;
XX Best Local Similarity 100.0%; Pred. No. 3.2e-58;
XX Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 82 LGDGTAPQKULFPSEKICLKWQOQTHRVGAGLQNGTCFANAALQCLTYTPPLANYMLS 141
XX
XX 1 LGDGTAPQKULFPSEKICLKWQOQTHRVGAGLQNGTCFANAALQCLTYTPPLANYMLS 60
XX
XX 142 HEHST 147
XX
XX 61 HEHST 66

RESULT 24
ABG40214
ID ABG40214 standard; peptide; 66 AA.
XX
AC ABG40214;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human peptide encoded by genome-derived single exon probe SEQ ID 29879.
XX
KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX
XX Homo sapiens.
XX
XX WO200186003-A2.
XX
XX 15-NOV-2001.
XX
XX 30-JAN-2001; 2001WO-US000665.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2002-114183/15.
XX
XX Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples.
XX
XX Claim 27; SEQ ID NO 29879; 634pp; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human lung comprising single exon nucleic acid probes having one of
XX 12614 nucleic acid sequences mentioned in the specification, or their
XX complements or the 12387 open reading frames derived from the 12614
XX probes. Also included are a microarray comprising the novel set of probes
XX; the novel set of probes which hybridise at high stringency to a nucleic
XX acid expressed in the human lung; measuring gene expression in a sample
XX derived from human lung, comprising (a) contacting the array with a
XX collection of detectably labeled nucleic acids derived from human lung
XX mRNA, and (b) measuring the label detectably bound to each probe of the
XX array; identifying exons in a eukaryotic genome, comprising (a)
XX algorithmically predicting at least one exon from genomic sequences of
XX the eukaryote; and (b) detecting specific hybridisation of detectably
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
XX having a fragment identical to the predicted exon, the probe is included
XX in the above mentioned microarray; assigning exons to a single gene,
XX comprising (a) identifying exons from genomic sequence by the method
XX above and (b) measuring the expression of each of the exons in several
XX tissues and/or cell types using hybridisation to a single exon
XX microarrays having a probe with the exon, where a common pattern of
XX expression of the exons in the tissues and/or cell types indicates that
XX the exons should be assigned to a single gene; a peptide comprising one
XX of 12011 sequences, mentioned in the specification, or encoded by the
XX probes/open reading frames (ORF). The probes are used for gene expression
XX analysis, and for identifying exons in a gene, particularly using human

CC lung derived mRNA and for the study of lung diseases such as asthma, lung
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
CC tuberculous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
CC present sequence is a peptide/protein encoded by a single exon probe of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 66 AA;

Query Match 8.5%; Score 66; DB 5; Length 66;
Best Local Similarity 100.0%; Pred. No. 3.2e-58;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 82 LGDGIAPPQKVPFSEKICLKWKQOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYMLS 141
Db 1 LGDGIAPPQKVPFSEKICLKWKQOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYMLS 60
QY 142 HEHSKT 147
Db 61 HEHSKT 66

RESULT 25
AAM84526
ID AAM84526 standard; protein; 133 AA.
AC AAM84526;
XX 07-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen SEQ ID NO:12119.
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
XX Homo sapiens.
XX WO200157182-A2.
XX 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US001354.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 18-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-023263P.
PR 14-SEP-2000; 2000US-023264P.
PR 14-SEP-2000; 2000US-023265P.
PR 21-SEP-2000; 2000US-0232223P.
PR 21-SEP-2000; 2000US-0232274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0245474P.
PR 08-NOV-2000; 2000US-0245475P.
PR 08-NOV-2000; 2000US-0245476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.

ADC27049
ID ADC27049 standard; protein; 248 AA.
XX
XX AC
XX ADC27049;
XX
DT 18-DEC-2003 (first entry)
XX
XX Human deubiquitinating enzyme DUB8.10 SEQ ID NO:46.
XX human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
XX infection; antiinflammatory; immunosuppressive; hDUB8; chromosome 8.
XX
XX Homo sapiens.
XX
XX WO2003072724-A2.
XX
XX 04-SEP-2003.
XX
XX 20-FEB-2003; 2003WO-US005338.
XX
XX 22-FEB-2002; 2002US-0358873P.
XX PR 22-FEB-2002; 2002US-0358875P.
XX PR 08-MAR-2002; 2002US-0363020P.
XX PR 12-APR-2002; 2002GB-00008404.
XX
XX (AVET) AVENTIS PHARM INC.
XX
XX Hahn C, Liu H;
XX
XX WFI; 2003-721759/68.
XX DR N-PSDB; ADC27050.
XX
XX New isolated polynucleotide encoding human deubiquitinating proteases,
XX useful for identifying e.g. antiinflammatory or immunosuppressive agents,
XX also the encoded polypeptides.
XX
XX Claim 2; SEQ ID NO 46; 201pp; English.
XX
XX The present invention describes human deubiquitinating protease enzymes
XX (I), designated hDUB. The enzymes designated hDUB4 are located on
XX chromosome 4, and the enzymes designated hDUB8 are located on chromosome
XX 8. Also described is a method for reducing inflammation, modulating
XX autoimmune diseases and modulating immune reactions during infections by
XX administering a compound (A) that inhibits (I) or alters regulation of
XX the transcription of nucleic acid sequences encoding (I). (I) has
XX antiinflammatory and immunosuppressive activities, and can be used in the
XX regulation of cytokine/cytokine receptor signaling involved in lymphocyte
XX proliferation. (I) can be useful for identifying their respective
XX inhibitors (A). (A) are useful for reducing inflammation and to modulate
XX autoimmune diseases or immune reactions during infections. The present
XX sequence represents hDUB8.10 from the present invention.
XX
XX Sequence 248 AA;
XX
XX Query Match 2.1%; Score 16; DB 7; Length 248;
XX Best Local Similarity 100.0%; Pred.No. 4.5e-07;
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 126 LQCLTYTPPLANYMLS 141
XX |||||||||||||
XX Db 95 LQCLTYTPPLANYMLS 110
XX
XX RESULT 28
XX ADC27047
XX ID ADC27047 standard; protein; 271 AA.
XX AC
XX ADC27047;
XX
XX DT 18-DEC-2003 (first entry)
XX
XX Human deubiquitinating enzyme DUB8.2 SEQ ID NO:44.
XX
XX

XX 06-NOV-2001; 2001US-0332633P.
 XX (DECO-) DECODE GENETICS EHF.
 PA Martinez RAM, Sigurdson GT;
 PI WPI; 2003-441582/41.
 XX N-PSDB; ADA50505.
 DR
 DR
 XX
 XX
 PT Novel isolated protease polypeptide and polynucleotide encoding the
 PT polypeptide useful for diagnosing and treating diseases or conditions
 PT associated with a protease.
 XX
 XX
 PS Claim 1; Page 66; 160pp; English.
 XX
 XX The invention relates to a novel isolated polypeptide comprising an amino
 CC acid sequence that has greater than 95 % identity to any one of 47 150-
 CC 350 residue protease polypeptide sequences, given in the specification.
 CC The nucleic acids, probes, primers, polypeptides and antibodies of the
 CC invention can be used in methods of diagnosis of a susceptibility to a
 CC disease or condition associated with a protease. The present sequence
 CC represents a protease of the invention.
 XX
 XX Sequence 308 AA;
 SQ
 Query Match 2.1%; Score 16; DB 6; Length 308;
 Best Local Similarity 100.0%; Pred. No. 5.5e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 126 LQCLTYTPPLANYMLS 141
 Db 26 LQCLTYTPPLANYMLS 41
 |||||
 RESULT 30
 AAEE19174
 ID AAEE19174 standard; protein; 314 AA.
 XX
 AC AAEE19174;
 XX
 DT 21-MAY-2002 (first entry)
 XX
 DE Human protease, PRTS-11 protein.
 XX
 KW Human; protease; PRTS-11; enzyme; gastritis; cirrhosis; Crohn's disease;
 KW gastrointestinal disorder; autoimmune; inflammatory; cell proliferative;
 KW cardiovascular; developmental; epithelial; neurological; reproductive;
 KW AIDS; Acquired Immune Deficiency Syndrome; allergy; rheumatoid arthritis;
 KW anaemia; asthma; atherosclerosis; hypertension; myocardial infarction;
 KW hepatitis; cancer; psoriasis; Cushing's syndrome; hypothyroidism; eczema;
 KW epilepsy; Alzheimer's disease; Huntington's disease; Parkinson's disease;
 KW Pick's disease; infertility; vitiligo; drug screening; gene therapy;
 KW Chromosome 16.
 XX
 OS Homo sapiens.
 XX
 PN WO200208396-A2.
 XX
 PD 31-JAN-2002.
 XX
 PF 17-JUL-2001; 2001WO-US022397.
 XX
 XX 21-JUL-2000; 2000US-0220063P.
 PR 28-JUL-2000; 2000US-0221680P.
 PR 04-AUG-2000; 2000US-0223544P.
 PR 11-AUG-2000; 2000US-0224717P.
 PR 16-AUG-2000; 2000US-0225988P.
 PR 23-AUG-2000; 2000US-0227568P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 XX Delegeane AM, Gandhi AR, Hafalia AJA, Lu DAM, Patterson C;

PI Tribouley CM, Das D, Kallick DA, Nguyen DB, Lee EA, Khan FA, Yue H;
 PI Au-Young J, Griffin JA, Policky JL, Ramkumar J, Yang J;
 PI Thangavelu K, Ding L, Kearney L, Baughn MR, Borowsky ML;
 PI Sanjanwala MS, Yao MG, Burford N, Walia NK, Lal P, Lee S, Todd S;
 PI Lo TP, Tang YT, Elliott VS, Azimzai Y, Lu Y;
 XX
 XX WPI; 2002-206082/26.
 DR N-PSDB; AAD30578.
 DR
 XX
 XX
 PT New human protease polypeptide, useful in diagnosis, prevention and
 PT treatment of gastrointestinal, cardiovascular, autoimmune/inflammatory,
 PT cell proliferative, developmental, epithelial and neurological disorders.
 XX
 XX Claim 1; Page 148; 182pp; English.
 XX
 XX The invention relates to an isolated human protease polypeptide (PRTS).
 CC PRTS protein and DNA are useful for diagnosing, treating and preventing
 CC gastrointestinal disorders (gastritis, cirrhosis, Crohn's disease),
 CC autoimmune/inflammatory disorders (AIDS, allergy, rheumatoid arthritis,
 CC anaemia, asthma), cardiovascular disorder (atherosclerosis, hypertension,
 CC myocardial infarction), cell proliferative disorders (hepatitis, cancer,
 CC psoriasis), developmental disorders (Cushing's syndrome, hypothyroidism),
 CC epithelial disorder (vitiligo, keloid, eczema), neurological disorders
 CC (epilepsy, Alzheimer's disease, Pick's disease, Huntington's disease,
 CC Parkinson's disease), and reproductive disorders (infertility). PRTS
 CC protein is useful in a number of drug screening techniques and to analyse
 CC the proteome of a tissue or cell type. PRTS DNA is useful for creating
 CC knockin humanised animals or transgenic animals to model human diseases,
 CC in somatic or germline gene therapy and in microarrays utilising fluids
 CC or tissues from patients to detect altered PKIN expression. The present
 CC sequence is human PRTS-11 protein
 XX
 XX Sequence 314 AA;
 SQ
 Query Match 2.1%; Score 16; DB 5; Length 314;
 Best Local Similarity 100.0%; Pred. No. 5.6e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 126 LQCLTYTPPLANYMLS 141
 Db 95 LQCLTYTPPLANYMLS 110
 |||||
 RESULT 31
 ADC27043
 ID ADC27043 standard; protein; 357 AA.
 XX
 AC ADC27043;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human deubiquitinating enzyme DUB4.4 SEQ ID NO:40.
 XX
 KW human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
 KW infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
 XX
 OS Homo sapiens.
 XX
 PN WO2003072724-A2.
 XX
 PD 04-SEP-2003.
 XX
 XX 20-FEB-2003; 2003WO-US005338.
 PF
 XX 22-FEB-2002; 2002US-0358873P.
 PR 22-FEB-2002; 2002US-0358875P.
 PR 08-MAR-2002; 2002US-0363020P.
 PR 12-APR-2002; 2002GB-00008404.
 XX
 PA (AVET) AVENTIS PHARM INC.
 XX
 XX Hahn C, Liu H;
 PI

DR WPI; 2003-721759/68.
DR N-PSDB; ADC27042.
XX
PT New isolated polynucleotide encoding human deubiquitinating proteases,
PT useful for identifying e.g. antiinflammatory or immunosuppressive agents,
PT also the encoded polypeptides.
XX
XX Claim 2; SEQ ID NO 40; 201pp; English.
XX
XX The present invention describes human deubiquitinating protease enzymes
CC (I), designated hDUB. The enzymes designated hDUB4 are located on
CC chromosome 4, and the enzymes designated hDUB8 are located on chromosome
CC 8. Also described is a method for reducing inflammation, modulating
CC autoimmune diseases and modulating immune reactions during infection of
CC administering a compound (A) that inhibits (I) or alters regulation of
CC the transcription of nucleic acid sequences encoding (I). (I) has
CC antiinflammatory and immunosuppressive activities, and can be used in the
CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
CC proliferation. (I) can be used for identifying their respective
CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
CC autoimmune diseases or immune reactions during infections. The present
CC sequence represents hDUB4.4 from the present invention.
XX
XX Sequence 357 AA;
SQ

Query Match 2.1%; Score 16; DB 7; Length 357;
Best Local Similarity 100.0%; Pred. No. 6.3e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 LQCLTYTPPLANYMIS 141
Db 95 LQCLTYTPPLANYMIS 110

RESULT 32
ADC27045
ID ADC27045 standard; protein; 398 AA.
XX
AC ADC27045;
XX
XX 18-DEC-2003 (first entry)
XX
XX Human deubiquitinating enzyme DUB4.9 SEQ ID NO:42.
XX
XX human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
XX infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
XX
XX Homo sapiens.
XX
XX WO2003072724-A2.
XX
XX 04-SEP-2003.
XX
XX 20-FEB-2003; 2003WO-US005338.
XX
XX 22-FEB-2002; 2002US-0358873P.
XX
XX 22-FEB-2002; 2002US-0358873P.
XX
XX 08-MAR-2002; 2002US-0363020P.
XX
XX 12-APR-2002; 2002GB-00008404.
XX
XX (AVET) AVENTIS PHARM INC.
XX
XX Hahn C, Liu H;
XX
XX WPI; 2003-721759/68.
XX
XX N-PSDB; ADC27044.
XX
XX New isolated polynucleotide encoding human deubiquitinating proteases,
XX useful for identifying e.g. antiinflammatory or immunosuppressive agents,
XX also the encoded polypeptides.
XX
XX Claim 2; SEQ ID NO 42; 201pp; English.
XX
XX

CC The present invention describes human deubiquitinating protease enzymes
CC (I), designated hDUB. The enzymes designated hDUB4 are located on
CC chromosome 4, and the enzymes designated hDUB8 are located on chromosome
CC 8. Also described is a method for reducing inflammation, modulating
CC autoimmune diseases and modulating immune reactions during infection of
CC administering a compound (A) that inhibits (I) or alters regulation of
CC the transcription of nucleic acid sequences encoding (I). (I) has
CC antiinflammatory and immunosuppressive activities, and can be used in the
CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
CC proliferation. (I) can be used for identifying their respective
CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
CC autoimmune diseases or immune reactions during infections. The present
CC sequence represents hDUB4.9 from the present invention.
XX
XX Sequence 398 AA;
SQ

Query Match 2.1%; Score 16; DB 7; Length 398;
Best Local Similarity 100.0%; Pred. No. 7e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 LQCLTYTPPLANYMIS 141
Db 95 LQCLTYTPPLANYMIS 110

RESULT 33
AAY77471
ID AAY77471 standard; protein; 530 AA.
XX
AC AAY77471;
XX
XX 05-JUN-2000 (first entry)
XX
XX Human deubiquitinating protein Dub11, SEQ ID NO:34.
XX
XX Human; deubiquitinating protein; Dub11; Dub12; immune disorder;
XX inflammation; allergy; immunosuppressant; antiarthritic; antirheumatoid;
XX antiinflammatory; dermatological; antithyroid.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Misc-difference 279 /label= Xaa
XX /note= 'Xaa is unknown'
XX
XX WO2000010197-A2.
XX
XX 13-JAN-2000.
XX
XX 06-JUL-1999; 99WO-US012366.
XX
XX 06-JUL-1998; 98US-00110938.
XX
XX 13-JUL-1998; 98US-00114466.
XX
XX 23-JUL-1998; 98US-0093897P.
XX
XX 12-AUG-1998; 98US-00132968.
XX
XX 18-AUG-1998; 98US-00136214.
XX
XX 11-SEP-1998; 98US-0093999P.
XX
XX (SCHE) SCHERING CORP.
XX
XX Bates EEM, Lebecque SJE, Murphy EE, Mattson JD, Gorman DM;
XX Hedrick JA, Wang L, Zlotnik A, Murgolo N, Greene JR, Johnston JA;
XX Bazan JF, Mahony D, Lees EM;
XX
XX WPI; 2000-171015/15.
XX
XX N-PSDB; AAZ92414.
XX
XX New isolated mammalian genes, used to develop products for treating e.g.
XX immune, inflammatory or allergic abnormalities, cancers or degenerative
XX conditions.
XX
XX Claim 43; Page 187-189; 218pp; English.
XX
XX

XX The invention relates to a number of primate and/or rodent proteins, and
 CC the genes which encode them. The invention encompasses human dendritic
 CC cell prostaglandin transporter (DC-PGT); the TNF (tumour necrosis factor)
 CC receptor family-related proteins HDPEA84, HSLJD37R and RANKL; human CC
 CC chemokine HCC5; human deubiquitinating proteins Dub11 and Dub 12; human
 CC MD-1 and human and murine MD-2 proteins, which exhibit the properties of
 CC ligands for proteins comprising a leucine-rich motif (LRR); human cyclin
 CC E2; cDNAs encoding these proteins; and antibodies against these proteins.
 CC The proteins can be used for modulating the physiology or development of
 CC a cell. They can be used to mediate uptake of substrates (e.g.,
 CC prostaglandin-like molecules), to modulate or mediate cellular
 CC interactions (e.g., induce or prevent trafficking, proliferation, or
 CC differentiation of cells), or are intracellular proteins which are
 CC important in various cellular processes such as the deubiquitination of
 CC proteins or cell cycle regulation. The products can be used for treating
 CC medical conditions such as immune, inflammatory or allergic disorders, or
 CC abnormal cellular proliferation, for example, cancers or degenerative
 CC conditions. They can be used to modulate immune responses in disease
 CC states e.g., autoimmune disorders, including rheumatoid arthritis, as
 CC systemic lupus erythematosus, Hashimoto's autoimmune thyroiditis, as well
 CC as acute and chronic inflammatory responses in which T cell activation,
 CC expansion, and/or immunological T cell memory play an important role.
 CC Sequences AAY77470-Y77473 represent human deubiquitinating proteins
 CC (Dub). AAY77470-Y77471 encode Dub11 proteins and AAY77472-Y77473 encode
 CC Dub12 proteins
 XX
 XX Sequence 530 AA;

Query Match 2.1%; Score 16; DB 3; Length 530;
 Best Local Similarity 100.0%; Pred. No. 9.1e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LOCLTYTPPLANYMLS 141
 DB 95 LOCLTYTPPLANYMLS 110

RESULT 34
 ABU61836
 ID ABU61836 standard; protein; 530 AA.

XX ABU61836;
 AC
 XX 15-AUG-2003 (first entry)
 XX Updated human de-ubiquitination protein Dub11.

XX Human; enzyme; de-ubiquitination; Dub11; cell development;
 KW tissue culture development.

XX Homo sapiens.

XX Key Location/Qualifiers
 FT Misc-difference 279
 FT /label= Unknown
 FT /note= "Encoded by GTM"

PN US2003028005-A1.

XX 06-FEB-2003.

XX 11-AUG-1999; 99US-00372348.

XX 12-AUG-1998; 98US-0096328P.

XX (BAZA/) BAZAN J F.

XX Bazan JF;

XX WPI; 2003-466159/44.

XX N-P5DB; ACA62458.

XX

PT New composition comprising HCC5, Dub11, Dub12, primate MD1 or MD2 or
 FT rodent MD2 polypeptide, useful for modulating physiology or development
 XX of a cell or tissue culture cells.
 PS Claim 1; Page 8-10; 77pp; English.
 XX The invention relates to a novel chemokine, de-ubiquitination or cell
 CC surface protein. The composition is useful for modulating physiology or
 CC development of a cell or tissue culture cells. The present sequence
 CC represents the amino acid sequence of the updated human de-ubiquitination
 CC protein Dub11
 XX
 XX Sequence 530 AA;

Query Match 2.1%; Score 16; DB 6; Length 530;
 Best Local Similarity 100.0%; Pred. No. 9.1e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LOCLTYTPPLANYMLS 141
 DB 95 LOCLTYTPPLANYMLS 110

RESULT 35
 ABU10207
 ID ABU10207 standard; protein; 530 AA.

XX ABU10207;

XX 28-JUL-2003 (first entry)

DE Human ubiquitin carboxyl-terminal hydrolase 68999.

XX Human; enzyme; ubiquitin carboxyl-terminal hydrolase; tissue typing;
 KW cell proliferation disorder; cell differentiation disorder; cancer;
 KW hormonal disorder; neurological disorder; Parkinson's disease; rickets;
 KW Alzheimer's disease; viral infection; bacterial infection; diabetes;
 KW parasitic infection; bone metabolism disorder; osteoporosis; arrhythmia;
 KW immune disorder; multiple sclerosis; cardiovascular disorder; psoriasis;
 KW congestive heart failure; blood vessel disorder; hypertension; psoriasis;
 KW Kawasaki syndrome; endothelial cell disorder; Grave's disease; obesity;
 KW liver disorder; hepatitis; Gaucher's disease; pain disorder;
 KW metabolic disorder.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Modified-site 22..24
 FT /note= "Protein kinase C phosphorylation site"

FT Modified-site 23..26
 FT /note= "Casein kinase II phosphorylation site"

FT Modified-site 36..39
 FT /note= "Casein kinase II phosphorylation site"

FT Modified-site 71..73
 FT /note= "Protein kinase C phosphorylation site"

FT Domain 80..111
 FT /label= UCH1

FT Domain 81..96
 FT /note= "Ubiquitin carboxyl-terminal hydrolase-1 domain"

FT Domain 81..96
 FT /note= "Ubiquitin carboxyl-terminal hydrolase family 2
 signature 1 domain"

FT Modified-site 122..127

FT Binding-site 207..212
 FT /note= "N-myristoylation site"

FT Modified-site 212..217
 FT /note= "Cytochrome c family haem binding site"

FT Domain 264..279
 FT /note= "N-myristoylation site"

FT Modified-site 272..274
 FT /note= "Helix-loop-helix dimerisation domain"

FT Modified-site 281..284
 FT /note= "Protein kinase C phosphorylation site"

FT Modified-site 281..284
 FT /note= "cAMP and cGMP-dependent protein kinase"

FT phosphorylation site"
FT 288..293
FT /note= "N-myristoylation site"
FT Domain
FT 313..374
FT /label= UCH2
FT /note= "Ubiquitin carboxyl-terminal hydrolase-2 domain"
FT Domain
FT 317..335
FT /note= "Ubiquitin carboxyl-terminal hydrolase family 2
FT signature 2 domain"
FT 340..347
FT /note= "Tyrosine kinase phosphorylation site"
FT Modified-site
FT 376..379
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site
FT 395..400
FT /note= "N-myristoylation site"
FT Modified-site
FT 399..401
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site
FT 401..404
FT /note= "cAMP and cGMP-dependent protein kinase
FT phosphorylation site"
FT 404..407
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site
FT 432..435
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site
FT 446..449
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site
FT 495..498
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site
FT 503..508
FT /note= "N-myristoylation site"
FT Modified-site
FT 512..514
FT /note= "Protein kinase C phosphorylation site"
US2003022201-A1.
30-JAN-2003.
27-MAR-2002; 2002US-00107695.
27-MAR-2001; 2001US-0279184P.
(MILL-) MILLENNIUM PHARM INC.
Kapeller-Libermann R;
WPI; 2003-447421/42.
N-PSDB; ACA61698.
New 68999 nucleic acid molecule, useful for treating, preventing and/or
diagnosing 68999-mediated disorders, e.g. cancer, diabetes, arrhythmia,
pain or obesity, and in chromosome mapping, tissue typing and forensic
biology.
Claim 13; Page 40-41; 49pp; English.
The invention relates to an isolated 68999 nucleic acid molecule which
encodes a ubiquitin carboxyl-terminal hydrolase. The nucleic acid
molecules, polypeptides and antibodies are useful in screening assays,
predictive medicine (e.g. diagnostic assays, prognostic assays,
monitoring clinical trials and pharmacogenetics) and prophylactic and
therapeutic methods. The nucleic acids may also be used in chromosome
mapping, tissue typing and forensic biology and as surrogate markers. The
nucleic acid and polypeptide may be used for treating, preventing and/or
diagnosing 68999-mediated or -related disorders, such as cell
proliferation and/or differentiation disorders (e.g. cancers), hormonal
disorders, neurological disorders (e.g. Parkinson's or Alzheimer's
disease), infections (e.g. viral, bacterial or parasitic), bone
metabolism disorders (e.g. osteoporosis and rickets), immune disorders
(e.g. diabetes and multiple sclerosis), cardiovascular disorders (e.g.
arrhythmias congestive heart failure), blood vessel disorders (e.g.
hypertension and Kawasaki syndrome), endothelial cell disorder (e.g.
psoriasis and Grave's disease) liver disorders (e.g. hepatitis and
Gaucher's disease) and pain or metabolic disorders (e.g. obesity or

CC cachexia). The present sequence represents the amino acid sequence of
CC human ubiquitin carboxyl-terminal hydrolase 68999
XX
SQ Sequence 530 AA;
Query Match 2.1%; Score 16; DB 6; Length 530;
Best Local Similarity 100.0%; Pred. No. 9.1e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 126 LQCLTVTPPLANYMLS 141
Db 95 LQCLTVTPPLANYMLS 110
RESULT 36
ADC27021
ID ADC27021 standard; protein; 530 AA.
XX
AC ADC27021;
XX 18-DEC-2003 (first entry)
DT Human deubiquitinating enzyme DUB4.7 SEQ ID NO:18.
XX human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
XX infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
XX Homo sapiens.
XX WO2003072724-A2.
XX 04-SEP-2003.
XX 20-FEB-2003; 2003WO-US005338.
XX 22-FEB-2002; 2002US-0358873P.
XX 22-FEB-2002; 2002US-0358875P.
PR 08-MAR-2002; 2002US-0361020P.
PR 12-APR-2002; 2002GB-00008404.
XX (AVET) AVENTIS PHARM INC.
XX Hahn C, Liu H;
XX WPI; 2003-721759/68.
DR N-PSDB; ADC27020.
XX
PT New isolated polynucleotide encoding human deubiquitinating proteases,
PT useful for identifying a.g. antiinflammatory or immunosuppressive agents,
PT also the encoded polypeptides.
XX Claim 2; SEQ ID NO 18; 201pp; English.
CC The present invention describes human deubiquitinating protease enzymes
CC (I), designated hDUB. The enzymes designated hDUB4 are located on
CC chromosome 4, and the enzymes designated hDUB8 are located on chromosome
CC 8. Also described is a method for reducing inflammation, modulating
CC autoimmune diseases and modulating immune reactions during infections by
CC administering a compound (A) that inhibits (I) or alters regulation of
CC the transcription of nucleic acid sequences encoding (I). (I) has
CC antiinflammatory and immunosuppressive activities, and can be used in the
CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
CC proliferation. (I) can be used for identifying their respective
CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
CC autoimmune diseases or immune reactions during infections. The present
CC sequence represents hDUB4.7 from the present invention.
XX
SQ Sequence 530 AA;
Query Match 2.1%; Score 16; DB 7; Length 530;
Best Local Similarity 100.0%; Pred. No. 9.1e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141
|||||
Db 95 LQCLTYTPPLANYMLS 110

RESULT 37
ADC27033
ID ADC27033 standard; protein; 530 AA.

AC ADC27033;
XX 18-DEC-2003 (first entry)
XX Human deubiquitinating enzyme DUB8.5 SEQ ID NO:30.

XX human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
KW infection; antiinflammatory; immunosuppressive; hDUB8; chromosome 8.

XX Homo sapiens.

OS WO2003072724-A2.

PN 12-APR-2002; 2002GB-00008404.

XX 04-SEP-2003.

XX 20-FEB-2003; 2003WO-US005338.

XX 22-FEB-2002; 2002US-0358873P.

PR 22-FEB-2002; 2002US-0358875P.

PR 08-MAR-2002; 2002US-0363020P.

PR 12-APR-2002; 2002GB-00008404.

XX (AVET) AVENTIS PHARM INC.

PA Hahn C, Liu H;

PI WPI; 2003-721759/68.

XX N-PSDB; ADC27032.

XX New isolated polynucleotide encoding human deubiquitinating proteases,
PT useful for identifying e.g. antiinflammatory or immunosuppressive agents,
PT also the encoded polypeptides.

XX Claim 2; SEQ ID NO 30; 201pp; English.

XX The present invention describes human deubiquitinating protease enzymes
(I), designated hDUB. The enzymes designated hDUB4 are located on
CC chromosome 4, and the enzymes designated hDUB8 are located on chromosome
CC 8. Also described is a method for reducing inflammation, modulating
CC autoimmune diseases and modulating immune reactions during infections by
CC administering a compound (A) that inhibits (I) or alters regulation of
CC the transcription of nucleic acid sequences encoding (I). (I) has
CC antiinflammatory and immunosuppressive activities, and can be used in the
CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
CC proliferation. (I) can be used for identifying their respective
CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
CC autoimmune diseases or immune reactions during infections. The present
CC sequence represents hDUB8.5 from the present invention.

XX Sequence 530 AA;

Query Match 2.1%; Score 16; DB 7; Length 530;
Best Local Similarity 100.0%; Pred. No. 9.1e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141
|||||
Db 95 LQCLTYTPPLANYMLS 110

RESULT 38
ADC27019
ID ADC27019 standard; protein; 530 AA.

AC ADC27019;

XX 18-DEC-2003 (first entry)

XX Human deubiquitinating enzyme DUB4.6 SEQ ID NO:16.

XX human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
KW infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.

XX Homo sapiens.

OS WO2003072724-A2.

PN 04-SEP-2003.

XX 20-FEB-2003; 2003WO-US005338.

XX 22-FEB-2002; 2002US-0358873P.

PR 22-FEB-2002; 2002US-0358875P.

PR 08-MAR-2002; 2002US-0363020P.

PR 12-APR-2002; 2002GB-00008404.

XX (AVET) AVENTIS PHARM INC.

PA Hahn C, Liu H;

PI WPI; 2003-721759/68.

XX N-PSDB; ADC27018.

XX New isolated polynucleotide encoding human deubiquitinating proteases,
PT useful for identifying e.g. antiinflammatory or immunosuppressive agents,
PT also the encoded polypeptides.

XX Claim 2; SEQ ID NO 16; 201pp; English.

XX The present invention describes human deubiquitinating protease enzymes
(I), designated hDUB. The enzymes designated hDUB4 are located on
CC chromosome 4, and the enzymes designated hDUB8 are located on chromosome
CC 8. Also described is a method for reducing inflammation, modulating
CC autoimmune diseases and modulating immune reactions during infections by
CC administering a compound (A) that inhibits (I) or alters regulation of
CC the transcription of nucleic acid sequences encoding (I). (I) has
CC antiinflammatory and immunosuppressive activities, and can be used in the
CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
CC proliferation. (I) can be used for identifying their respective
CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
CC autoimmune diseases or immune reactions during infections. The present
CC sequence represents hDUB4.6 from the present invention.

XX Sequence 530 AA;

Query Match 2.1%; Score 16; DB 7; Length 530;
Best Local Similarity 100.0%; Pred. No. 9.1e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141
|||||
Db 95 LQCLTYTPPLANYMLS 110

RESULT 39
ADC27037
ID ADC27037 standard; protein; 530 AA.

XX ADC27037;

XX 18-DEC-2003 (first entry)

XX Human deubiquitinating enzyme DUB8.7 SEQ ID NO:34.

XX human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
KW infection; antiinflammatory; immunosuppressive; hDUB8; chromosome 8.

OS Homo sapiens.
 XX W02003072724-A2.
 XX
 XX PD 04-SEP-2003.
 XX
 XX PF 20-FEB-2003; 2003WO-US005338.
 XX
 XX PR 22-FEB-2002; 2002US-0358873P.
 XX PR 22-FEB-2002; 2002US-0358873P.
 XX PR 08-MAR-2002; 2002US-0363020P.
 XX PR 12-APR-2002; 2002GB-00008404.
 XX
 XX PA (AVET) AVENTIS PHARM INC.
 XX
 XX PI Hahn C, Liu H;
 XX
 XX WPI; 2003-721759/68.
 XX N-PSDB; ADC27036.
 XX
 XX New isolated polynucleotide encoding human deubiquitinating proteases,
 XX useful for identifying e.g. antiinflammatory or immunosuppressive agents,
 XX PT also the encoded polypeptides.
 XX
 XX PS Claim 2; SEQ ID NO 34; 201pp; English.
 XX
 XX The present invention describes human deubiquitinating protease enzymes
 CC (1), designated hDUB. The enzymes designated hDUB4 are located on
 CC chromosome 4, and the enzymes designated hDUB8 are located on chromosome
 CC 8. Also described is a method for reducing inflammation, modulating
 CC autoimmune diseases and modulating immune reactions during infections by
 CC administering a compound (A) that inhibits (I) or alters regulation of
 CC the transcription of nucleic acid sequences encoding (I). (I) has
 CC antiinflammatory and immunosuppressive activities, and can be used in the
 CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
 CC proliferation. (I) can be used for identifying their respective
 CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
 CC autoimmune diseases or immune reactions during infections. The present
 CC sequence represents hDUB8.7 from the present invention.
 XX
 XX SQ Sequence 530 AA;
 XX
 XX Query Match 2.1%; Score 16; DB 7; Length 530;
 XX Best Local Similarity 100.0%; Pred. No. 9.1e-07;
 XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX QY 126 LQCLTYTPPLANYMLS 141
 XX |||||
 XX 95 LQCLTYTPPLANYMLS 110
 XX
 XX RESULT 40
 XX ADC27015
 XX ID ADC27015 standard; protein; 530 AA.
 XX
 XX AC ADC27015;
 XX
 XX XX 18-DEC-2003 (first entry)
 XX
 XX DE Human deubiquitinating enzyme DUB4.3 SEQ ID NO:12.
 XX
 XX KW human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
 XX infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
 XX
 XX OS Homo sapiens.
 XX
 XX XX W02003072724-A2.
 XX
 XX XX PD 04-SEP-2003.
 XX
 XX XX PF 20-FEB-2003; 2003WO-US005338.
 XX
 XX XX PR 22-FEB-2002; 2002US-0358873P.
 XX
 XX XX

PR 22-FEB-2002; 2002US-0358873P.
 PR 08-MAR-2002; 2002US-0363020P.
 PR 12-APR-2002; 2002GB-00008404.
 XX
 XX PA (AVET) AVENTIS PHARM INC.
 XX
 XX PI Hahn C, Liu H;
 XX
 XX WPI; 2003-721759/68.
 XX N-PSDB; ADC27014.
 XX
 XX New isolated polynucleotide encoding human deubiquitinating proteases,
 XX useful for identifying e.g. antiinflammatory or immunosuppressive agents,
 XX PT also the encoded polypeptides.
 XX
 XX PS Claim 2; SEQ ID NO 12; 201pp; English.
 XX
 XX The present invention describes human deubiquitinating protease enzymes
 CC (1), designated hDUB. The enzymes designated hDUB4 are located on
 CC chromosome 4, and the enzymes designated hDUB8 are located on chromosome
 CC 8. Also described is a method for reducing inflammation, modulating
 CC autoimmune diseases and modulating immune reactions during infections by
 CC administering a compound (A) that inhibits (I) or alters regulation of
 CC the transcription of nucleic acid sequences encoding (I). (I) has
 CC antiinflammatory and immunosuppressive activities, and can be used in the
 CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
 CC proliferation. (I) can be used for identifying their respective
 CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
 CC autoimmune diseases or immune reactions during infections. The present
 CC sequence represents hDUB4.3 from the present invention.
 XX
 XX SQ Sequence 530 AA;
 XX
 XX Query Match 2.1%; Score 16; DB 7; Length 530;
 XX Best Local Similarity 100.0%; Pred. No. 9.1e-07;
 XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX QY 126 LQCLTYTPPLANYMLS 141
 XX |||||
 XX 95 LQCLTYTPPLANYMLS 110
 XX
 XX RESULT 41
 XX ADC27023
 XX ID ADC27023 standard; protein; 565 AA.
 XX
 XX AC ADC27023;
 XX
 XX XX 18-DEC-2003 (first entry)
 XX
 XX DE Human deubiquitinating enzyme DUB4.8 SEQ ID NO:20.
 XX
 XX KW human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
 XX infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
 XX
 XX OS Homo sapiens.
 XX
 XX XX W02003072724-A2.
 XX
 XX XX PD 04-SEP-2003.
 XX
 XX XX PF 20-FEB-2003; 2003WO-US005338.
 XX
 XX XX PR 22-FEB-2002; 2002US-0358873P.
 XX PR 22-FEB-2002; 2002US-0358873P.
 XX PR 08-MAR-2002; 2002US-0363020P.
 XX PR 12-APR-2002; 2002GB-00008404.
 XX
 XX PA (AVET) AVENTIS PHARM INC.
 XX
 XX PI Hahn C, Liu H;
 XX
 XX WPI; 2003-721759/68.
 XX

DR N-PSDB; ADC27022.
 XX New isolated polynucleotide encoding human deubiquitinating proteases,
 PT useful for identifying e.g. antiinflammatory or immunosuppressive agents,
 PT also the encoded polypeptides.
 XX
 PS Claim 2; SEQ ID NO 20; 201pp; English.
 XX
 CC The present invention describes human deubiquitinating protease enzymes
 CC (I), designated hDUB. The enzymes designated hDUB4 are located on
 CC chromosome 4, and the enzymes designated hDUB8 are located on
 CC chromosome 8. Also described is a method for reducing inflammation, modulating
 CC autoimmune diseases and modulating immune reactions during infections by
 CC administering a compound (A) that inhibits (I) or alters regulation of
 CC the transcription of nucleic acid sequences encoding (I). (I) has
 CC antiinflammatory and immunosuppressive activities, and can be used in the
 CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
 CC proliferation. (I) can be used for identifying their respective
 CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
 CC autoimmune diseases or immune reactions during infections. The present
 CC sequence represents hDUB4.8 from the present invention.
 XX
 SQ Sequence 565 AA;

Query Match 2.1%; Score 16; DB 7; Length 565;
 Best Local Similarity 100.0%; Pred. No. 9.6e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141
 |||||
 Db 130 LQCLTYTPPLANYMLS 145

RESULT 42
 ADC27017
 ID ADC27017 standard; protein; 574 AA.
 XX
 AC ADC27017;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human deubiquitinating enzyme DUB4.5 SEQ ID NO:14.
 XX
 KW human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
 KW infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
 XX
 OS Homo sapiens.
 XX
 PN WO2003072724-A2.
 XX
 PD 04-SEP-2003.
 XX
 PF 20-FEB-2003; 2003WO-US005338.
 XX
 PR 22-FEB-2002; 2002US-0358873P.
 PR 22-FEB-2002; 2002US-0358875P.
 PR 08-MAR-2002; 2002US-0363020P.
 PR 12-APR-2002; 2002GB-00008404.
 XX
 PA (AVET) AVENTIS PHARM INC.
 XX
 PI Hahn C, Liu H;
 XX
 DR WPI; 2003-721759/68.
 DR N-PSDB; ADC27016.
 XX
 PT New isolated polynucleotide encoding human deubiquitinating proteases
 PT useful for identifying e.g. antiinflammatory or immunosuppressive agents,
 PT also the encoded polypeptides.
 XX
 PS Claim 2; SEQ ID NO 14; 201pp; English.
 XX
 CC The present invention describes human deubiquitinating protease enzymes

CC (I), designated hDUB. The enzymes designated hDUB4 are located on
 CC chromosome 4, and the enzymes designated hDUB8 are located on chromosome
 CC 8. Also described is a method for reducing inflammation, modulating
 CC autoimmune diseases and modulating immune reactions during infections by
 CC administering a compound (A) that inhibits (I) or alters regulation of
 CC the transcription of nucleic acid sequences encoding (I). (I) has
 CC antiinflammatory and immunosuppressive activities, and can be used in the
 CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
 CC proliferation. (I) can be used for identifying their respective
 CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
 CC autoimmune diseases or immune reactions during infections. The present
 CC sequence represents hDUB4.5 from the present invention.
 XX
 SQ Sequence 574 AA;

Query Match 2.1%; Score 16; DB 7; Length 574;
 Best Local Similarity 100.0%; Pred. No. 9.8e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141
 |||||
 Db 130 LQCLTYTPPLANYMLS 145

RESULT 43
 ADC27027
 ID ADC27027 standard; protein; 633 AA.
 XX
 AC ADC27027;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human deubiquitinating enzyme DUB4.11 SEQ ID NO:24.
 XX
 KW human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
 KW infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
 XX
 OS Homo sapiens.
 XX
 PN WO2003072724-A2.
 XX
 PD 04-SEP-2003.
 XX
 PF 20-FEB-2003; 2003WO-US005338.
 XX
 PR 22-FEB-2002; 2002US-0358873P.
 PR 22-FEB-2002; 2002US-0358875P.
 PR 08-MAR-2002; 2002US-0363020P.
 PR 12-APR-2002; 2002GB-00008404.
 XX
 PA (AVET) AVENTIS PHARM INC.
 XX
 PI Hahn C, Liu H;
 XX
 DR WPI; 2003-721759/68.
 DR N-PSDB; ADC27026.
 XX
 PT New isolated polynucleotide encoding human deubiquitinating proteases,
 PT useful for identifying e.g. antiinflammatory or immunosuppressive agents,
 PT also the encoded polypeptides.
 XX
 PS Claim 2; SEQ ID NO 24; 201pp; English.

The present invention describes human deubiquitinating protease enzymes
 CC (I), designated hDUB. The enzymes designated hDUB4 are located on
 CC chromosome 4, and the enzymes designated hDUB8 are located on
 CC chromosome 8. Also described is a method for reducing inflammation, modulating
 CC autoimmune diseases and modulating immune reactions during infections by
 CC administering a compound (A) that inhibits (I) or alters regulation of
 CC the transcription of nucleic acid sequences encoding (I). (I) has
 CC antiinflammatory and immunosuppressive activities, and can be used in the
 CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
 CC proliferation. (I) can be used for identifying their respective

CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
CC autoimmune diseases or immune reactions during infections. The present
CC sequence represents hDUB4.11 from the present invention.

XX SQ Sequence 633 AA;

Query Match 2.1%; Score 16; DB 7; Length 633;
Best Local Similarity 100.0%; Pred. No. 1.7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 126 LQCLTYTPPLANYMLS 141
| | | | | | | | | | | | | | | | | | | | | |
Db 198 LQCLTYTPPLANYMLS 213
| | | | | | | | | | | | | | | | | | | | | |

RESULT 44
AD27011
ID ADC27011 standard; protein; 1021 AA.

XX AC ADC27011;

XX DT 18-DEC-2003 (first entry)

XX DE Human deubiquitinating enzyme DUB4.2a SEQ ID NO:8.

XX KW human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
XX infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.

XX OS Homo sapiens.

XX FN WO2003072724-A2.

XX PD 04-SEP-2003.

XX PF 20-FEB-2003; 2003WO-US005338.

XX PR 22-FEB-2002; 2002US-0358873P.

XX PR 22-FEB-2002; 2002US-0358873P.

XX PR 08-MAR-2002; 2002US-0363020P.

XX PR 12-APR-2002; 2002GB-00008404.

XX PA (AVET) AVENTIS PHARM INC.

XX PI Hahn C, Liu H;

XX DR WPI; 2003-721759/68.

XX DR N-PSDB; ADC27010.

XX PT New isolated polynucleotide encoding human deubiquitinating proteases,
XX useful for identifying e.g. antiinflammatory or immunosuppressive agents,
XX also the encoded polypeptides.

XX PS Claim 2; SEQ ID NO 8; 201pp; English.

XX CC The present invention describes human deubiquitinating protease enzymes
XX (I), designated hDUB. The enzymes designated hDUB4 are located on
XX chromosome 4, and the enzymes designated hDUB8 are located on chromosome
XX 8. Also described is a method for reducing inflammation, modulating
XX autoimmune diseases and modulating immune reactions during infections by
XX administering a compound (A) that inhibits (I) or alters regulation of
XX the transcription of nucleic acid sequences encoding (I). (I) has
XX antiinflammatory and immunosuppressive activities, and can be used in the
XX regulation of cytokine/cytokine receptor signaling involved in lymphocyte
XX proliferation. (I) can be used for identifying their respective
XX inhibitors (A). (A) are useful for reducing inflammation and to modulate
XX autoimmune diseases or immune reactions during infections. The present
XX sequence represents hDUB4.2a from the present invention.

XX SQ Sequence 1021 AA;

Query Match 2.1%; Score 16; DB 7; Length 1021;
Best Local Similarity 100.0%; Pred. No. 1.7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 126 LQCLTYTPPLANYMLS 141
| | | | | | | | | | | | | | | | | | | | | |
Db 11 LQCLTYTPPLANYMLS 26
| | | | | | | | | | | | | | | | | | | | | |

RESULT 45

ADC31203

ID ADC31203 standard; protein; 497 AA.

XX AC ADC31203;

XX DT 18-DEC-2003 (first entry)

XX DE Human novel polypeptide sequence, SEQ ID NO:1285.

XX KW Human; diagnostic; drug screening; forensics; gene mapping;
XX biodiversity assessment; Parkinson's disease; Alzheimer's disease;
XX neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
XX ulcers; osteoporosis; autoimmune disease; cancer;
XX molecular weight marker; food supplement; antiparkinsonian; nootropic;
XX neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;
XX anticulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
XX gene therapy; chromosome 8p23.

XX OS Homo sapiens.

XX FN WO2003029271-A2.

XX PD 10-APR-2003.

XX PF 24-SEP-2002; 2002WO-US030474.

XX PR 24-SEP-2001; 2001US-0324631P.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;

XX PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;

XX PI Haley-Vicente D, Drmanac RT;

XX DR WPI; 2003-371981/35.

XX DR N-PSDB; ADC30232.

XX PT New polynucleotide and polypeptide useful for diagnosing, preventing or
XX treating conditions such as neurodegenerative diseases, anemias, platelet
XX disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
XX cancer.

XX PS Claim 20; SEQ ID NO 1285; 1185pp; English.

XX CC The invention relates to 971 novel human cDNA sequences (ADC29919-
XX ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
XX invention also relates to nucleic acid sequences over 99% identical with
XX the novel human cDNAs. The invention additionally encompasses expression
XX vectors and host cells comprising a nucleic acid of the invention; the
XX recombinant production of a polypeptide of the invention; an antibody
XX against a polypeptide of the invention; a method of detecting
XX polynucleotides or polypeptides of the invention; and methods of
XX identifying a compound which binds to a polypeptide of the invention. The
XX invention further discloses methods of preventing, treating or
XX ameliorating a medical condition; kits comprising polynucleotide probes
XX and/or monoclonal antibodies for carrying out the methods of the
XX invention; methods for the identification of compounds that modulate the
XX expression or activity of the polynucleotide and/or polypeptide; and 767
XX contig sequences corresponding to the cDNA sequences of the invention
XX (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
XX -ADC33394). The nucleic acids and polypeptides of the invention are
XX useful in diagnostics, drug screening, forensics, gene mapping, in the
XX identification of mutations responsible for genetic disorders or other
XX traits, for assessing biodiversity, and in producing many other types of
XX data and products dependent on DNA and amino acid sequences. They are
XX also used for treating diseases such as Parkinson's disease, Alzheimer's

CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a specifically
CC claimed human polypeptide sequence of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 497 AA;

Query Match 1.9%; Score 15; DB 7; Length 497;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 QCLTYTPPLANYMLS 141
Db 96 QCLTYTPPLANYMLS 110
|||||

RESULT 46
ADC27031
ID ADC27031 standard; protein; 497 AA.

XX AC ADC27031;
XX
XX
XX DT 18-DEC-2003 (first entry)
XX DE Human deubiquitinating enzyme DUB8.3 SEQ ID NO:28.

XX KW human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
XX infection; antiinflammatory; immunosuppressive; hDUB8; chromosome 8.

XX OS Homo sapiens.
XX
XX PN WO2003072724-A2.
XX PD 04-SEP-2003.

XX PF 20-FEB-2003; 2003WO-US005338.
XX PR 22-FEB-2002; 2002US-0358873P.
XX PR 22-FEB-2002; 2002US-0358875P.
XX PR 08-MAR-2002; 2002US-0363020P.
XX PR 12-APR-2002; 2002GB-00008404.

XX PA (AVET) AVENTIS PHARM INC.

XX PI Hahn C, Liu H;

XX DR WPI; 2003-721759/68.
XX DR N-PSDB; ADC27030.

XX PT New isolated polynucleotide encoding human deubiquitinating proteases,
XX useful for identifying e.g. antiinflammatory or immunosuppressive agents,
XX PT also the encoded polypeptides.

XX PS Claim 2; SEQ ID NO 28; 201pp; English.

XX CC The present invention describes human deubiquitinating protease enzymes
CC (1), designated hDUB. The enzymes designated hDUB4 are located on
CC chromosome 4, and the enzymes designated hDUB8 are located on chromosome
CC 8. Also described is a method for reducing inflammation, modulating
CC autoimmune diseases and modulating immune reactions during infections by
CC administering a compound (A) that inhibits (1) or alters regulation of
CC the transcription of nucleic acid sequences encoding (1). (1) has
CC antiinflammatory and immunosuppressive activities, and can be used in the
CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
CC proliferation. (1) can be used for identifying their respective
CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
CC autoimmune diseases or immune reactions during infections. The present

CC sequence represents hDUB8.3 from the present invention.

XX Sequence 497 AA;

Query Match 1.9%; Score 15; DB 7; Length 497;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 QCLTYTPPLANYMLS 141
Db 96 QCLTYTPPLANYMLS 110
|||||

RESULT 47
AAAY77472
ID AAAY77472 standard; protein; 245 AA.

XX AC AAAY77472;
XX
XX DT 05-JUN-2000 (first entry)

XX DE Human deubiquitinating protein Dub12.

XX KW Human; deubiquitinating protein; Dub11; Dub12; immune disorder;
XX inflammation; allergy; immunosuppressant; antiarthritic; antirheumatoid;
XX antiinflammatory; dermatological; antithyroid.

XX OS Homo sapiens.

XX PN WO200001817-A2.

XX PD 13-JAN-2000.

XX PF 06-JUL-1999; 99WO-US012366.

XX PR 06-JUL-1998; 98US-00110938.
XX PR 13-JUL-1998; 98US-00114466.
XX PR 23-JUL-1998; 98US-0093897P.
XX PR 12-AUG-1998; 98US-00132968.
XX PR 18-AUG-1998; 98US-00136214.
XX PR 11-SEP-1998; 98US-0099999P.

XX (SCHE) SCHERING CORP.

XX PI Bates EEM, Lebecque SJE, Murphy BE, Mattson JD, Gorman DM;

XX PI Hedrick JA, Wang L, Zlotnik A, Murgolo NJ, Greene JR, Johnston JA;
XX PI Bazan JF, Mahony D, Lees EM;

XX DR WPI; 2000-171015/15.
XX DR N-PSDB; AAZ92415.

XX PT New isolated mammalian genes, used to develop products for treating e.g.
XX immune, inflammatory or allergic abnormalities, cancers or degenerative
XX PT conditions.

XX PS Claim 43; Page 191-192; 218pp; English.

XX CC The invention relates to a number of primate and/or rodent proteins, and
XX the genes which encode them. The invention encompasses human dendritic
XX cell prostaglandin transporter (DC-PGT); the TNF (tumour necrosis factor)
XX receptor family-related proteins HPTA84, HSLJD37R and RANKL; human CC
XX chemokine HCC5; human deubiquitinating proteins Dub1 and Dub 12; human
XX MD-1 and human and murine MD-2 proteins, which exhibit the properties of
XX ligands for proteins comprising a leucine-rich motif (LRR); human cyclin
XX E2; cDNAs encoding these proteins; and antibodies against these proteins.
XX The proteins can be used for modulating the physiology or development of
XX a cell. They can be used to mediate uptake of substrates (e.g.,
XX prostaglandin-like molecules), to modulate or mediate cellular
XX interactions (e.g., induce or prevent trafficking, proliferation, or
XX differentiation of cells), or are intracellular proteins which are
XX important in various cellular processes such as the deubiquitination of
XX proteins or cell cycle regulation. The products can be used for treating
XX medical conditions such as immune, inflammatory or allergic disorders, or

CC abnormal cellular proliferation, for example, cancers or degenerative
 CC conditions. They can be used to modulate immune responses in disease
 CC states e.g., autoimmune disorders, including rheumatoid arthritis,
 CC systemic lupus erythematosus, Hashimoto's autoimmune thyroiditis, as well
 CC as acute and chronic inflammatory responses in which T cell activation,
 CC expansion, and/or immunological T cell memory play an important role.
 CC Sequences AAY77470-Y77473 represent human deubiquitinating proteins
 CC (Dub). AAY77470-Y77471 encode Dub1 proteins and AAY77472-Y77473 encode
 CC Dub12 proteins
 XX
 SQ Sequence 245 AA;

Query Match 1.8%; Score 14; DB 3; Length 245;
 Best Local Similarity 100.0%; Pred. No. 4.8e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 SNVLTSLKRFANF 323
 |||||
 Db 96 SNVLTSLKRFANF 109

RESULT 48

ABU61837
 ID ABU61837 standard; protein; 245 AA.

AC ABU61837;

DT 15-AUG-2003 (first entry)

DE Human de-ubiquitination protein Dub12.

KW Human; enzyme; de-ubiquitination; Dub12; cell development;
 KW tissue culture development.

OS Homo sapiens.

PH Key Location/Qualifiers

FT Misc-difference 66
 /label= Unknown

FT /note= "Encoded by GMA"

FT Misc-difference 197
 /label= Unknown

FT /note= "Encoded by TNA"

FT Misc-difference 222
 /label= Unknown

FT /note= "Encoded by NNN"

FT Misc-difference 236
 /label= Unknown

FT /note= "Encoded by AGN"

FT Misc-difference 239
 /label= Unknown

FT /note= "Encoded by NTC"

US2003028005-A1.

PD 06-FEB-2003.

PF 11-AUG-1999; 99US-00372348.

PR 12-AUG-1998; 98US-0096328P.

PA (BAZA/) BAZAN J F.

PI Bazan JF;

XX WPI; 2003-466159/44.

DR N-PSDB; ACA62459.

XX New composition comprising HCC5, Dub11, Dub12, primate MD1 or MD2 or
 PT rodent MD2 polypeptide, useful for modulating physiology or development
 PT of a cell or tissue culture cells.
 XX
 PS Claim 1; Page 10-11; 77pp; English.

XX
 CC The invention relates to a novel chemokine, de-ubiquitination or cell
 CC surface protein. The composition is useful for modulating physiology or
 CC development of a cell or tissue culture cells. The present sequence
 CC represents the amino acid sequence of the human de-ubiquitination protein
 CC Dub12
 XX
 SQ Sequence 245 AA;

Query Match 1.8%; Score 14; DB 6; Length 245;
 Best Local Similarity 100.0%; Pred. No. 4.8e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 SNVLTSLKRFANF 323
 |||||
 Db 96 SNVLTSLKRFANF 109

RESULT 49

AAB95208

ID AAB95208 standard; protein; 548 AA.

AC AAB95208;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:17313.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-0030253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00133767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.

XX Claim 8; SEQ ID NO 17313; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX
 SQ Sequence 548 AA;

Query Match 1.8%; Score 14; DB 4; Length 548;
 Best Local Similarity 100.0%; Pred. No. 9.9e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 SNVLTLSLKRFPANF 323
 |||||
 Db 321 SNVLTLSLKRFPANF 334

RESULT 50
 AAB95175
 ID AAB95175 standard; protein; 726 AA.
 XX
 AC AAB95175;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:17235.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-00116126.
 XX
 PR 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241699.
 XX
 PA (HELI-) HELIX RES INST.

XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Orsuki T;
 XX
 DR WPI; 2001-318749/34.

XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX
 PS Claim 8; SEQ ID NO 17235; 2537pp + Sequence Listing; English.

XX
 CC The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
 XX
 SQ Sequence 726 AA;

Query Match 1.8%; Score 14; DB 4; Length 726;
 Best Local Similarity 100.0%; Pred. No. 0.00013;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 SNVLTLSLKRFPANF 323
 |||||
 Db 21 SNVLTLSLKRFPANF 34

Search completed: August 10, 2004, 20:15:19
 Job time : 61 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:54:23 ; Search time 58 Seconds
(without alignments)
3775.419 Million cell updates/sec

Title: US-10-049-745-4

Perfect score: 4070

Sequence: 1 MTIVDKASESSDPAYQNQP.....AAESLEPPDAASLFFPSEG 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_29Jan04:*

1: geneseqp1980s:*

2: geneseqp1990s:*

3: geneseqp2000s:*

4: geneseqp2001s:*

5: geneseqp2002s:*

6: geneseqp2003as:*

7: geneseqp2003bs:*

8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4070	100.0	775	4 AAB74671	Aab74671 Human pro
2	4028	99.0	1251	5 AAU98889	Aau98889 Human pro
3	4022	98.8	1197	4 AAB95146	Aab95146 Human pro
4	4021	98.8	1270	5 AABP69291	Abp69291 Human pol
5	3999	98.3	762	4 AAB82177	Aab82177 Human ubi
6	3999	98.3	762	6 ABU08951	Abu08951 Human ubi
7	2908	71.4	558	5 ABB06118	Abb06118 Human NS
8	1423.5	35.0	1087	7 ADE31066	Ade31066 Human dia
9	1423.5	35.0	1123	4 AAY71889	Aay71889 Human ubi
10	1422.5	35.0	954	4 AAB93200	Aab93200 Human ubi
11	1375	33.8	548	4 AAB95208	Aab95208 Human pro
12	1150.5	28.3	565	7 ADC27023	Adc27023 Human pro
13	1148	28.2	530	6 ABU10207	Abu10207 Human ubi
14	1142.5	28.1	530	7 ADC27019	Adc27019 Human ubi
15	1140.5	28.0	530	7 ADC27015	Adc27015 Human pro
16	1139.5	28.0	574	7 ADC27017	Adc27017 Human pro
17	1135.5	27.9	530	7 ADC27021	Adc27021 Human pro
18	1132.5	27.8	715	7 ADC27025	Adc27025 Human pro
19	1128.5	27.7	1021	7 ADC27011	Adc27011 Human pro
20	1128	27.7	530	4 AAG64049	Aag64049 Human pro
21	1127	27.7	494	2 AAW30711	Aaw30711 Human ubi
22	1126.5	27.7	633	7 ADC27027	Adc27027 Human pro
23	1121.5	27.6	1016	7 ADC27007	Adc27007 Human pro
24	1118.5	27.5	530	7 ADC27013	Adc27013 Human pro
25	1114	27.4	398	7 ADC27045	Adc27045 Human pro

26	1113.5	27.4	530	3 AAY77471	Aay77471 Human deu
27	1113.5	27.4	530	6 ABU61836	Abu61836 Updated h
28	1111.5	27.3	530	7 ADC27009	Adc27009 Human deu
29	1096	26.9	530	7 ADC27037	Adc27037 Human deu
30	1073	26.4	530	7 ADC27029	Adc27029 Human deu
31	1052	25.8	530	6 ABR43250	Abr43250 Human PMM
32	1052	25.8	530	7 ADC27035	Adc27035 Human deu
33	1049.5	25.8	530	7 ADC27039	Adc27039 Human deu
34	1037	25.5	357	7 ADC27043	Adc27043 Human deu
35	1036	25.5	308	6 ADA50458	Ada50458 Human pro
36	1030	25.3	530	7 ADC27033	Adc27033 Human pro
37	999	24.5	308	6 ADA50460	Ada50460 Human pro
38	980.5	24.1	497	7 ADC31203	Adc31203 Human nov
39	980.5	24.1	497	7 ADC27031	Adc27031 Human nov
40	914	22.5	748	3 AAY77473	Aay77473 Human deu
41	914	22.5	748	6 ABU61838	Abu61838 Updated h
42	889.5	21.9	545	2 AAW30708	Aaw30708 Murine ub
43	888	21.8	545	3 AAY77475	Aay77475 Primate p
44	888	21.8	545	6 ABU61840	Abu61840 Human de-
45	846.5	20.8	521	2 AAW30710	Aaw30710 Murine ub

ALIGNMENTS

RESULT 1
AAB74671

ID AAB74671 standard; protein; 775 AA.

XX AAB74671;

DT 12-JUN-2001 (first entry)

DE Human protease and protease inhibitor PPTM-4.

KW Human; protease; protease inhibitor; protease and protease inhibitor;
KW PPTM; identification; diagnosis; anti-human immunodeficiency virus; HIV;
KW antidiabetic; immunostimulant; immunomodulator; antiinflammatory;
KW antithyroid; immunosuppressive; nephrotropic; antitumor; thyromimetic;
KW cytostatic; antibacterial; fungicide; protozoicide; antiarteriosclerotic;
KW antiatherosclerotic; antiparasitic; virucide; hepatotropic; gene therapy;
KW autoimmune disorder; inflammatory disease; SCID; Chediak-Higashi syndrome;
KW severe combined immunodeficiency disease; AIDS; Chediak-Higashi syndrome;
KW Cushing's disease; Addison's disease; Good pasture's syndrome; gout;
KW Crohn's disease; diabetes mellitus; Good pasture's syndrome; infection;
KW Grave's disease; Hashimoto's thyroiditis; Sjogren's syndrome; cancer;
KW Werner's syndrome; cell proliferative disorder; arteriosclerosis;
KW atherosclerosis; cirrhosis; hepatitis; psoriasis.

XX Homo sapiens.

WO200110903-A2.

15-FEB-2001.

09-AUG-2000; 2000WO-US021878.

09-AUG-1999; 99US-0147986P.

21-OCT-1999; 99US-0160807P.

(INCY-) INCYTE GENOMICS INC.

Yue H, Lal P, Tang YT, Bandman O, Baughn MR, Azimzai Y, Lu DAM;
Yang J;
WPI; 2001-202760/20.

N-PSDB; AAF81717.

New protease (inhibitors) useful for diagnosis and treatment of
autoimmune/inflammatory disorders such as acquired immunodeficiency
syndrome, Cushing's disease, Addison's disease and cell proliferative
disorders such as cancer.

Claim 1; Page 95-96; 13app; English.

AAAF81714 to AAF81740 encode the human proteases and protease inhibitors (PPIMs) given in AAB74668 to AAB74694. The PPIMs can have activities such as: anti-human immunodeficiency virus (HIV); antidiabetic; antichytoid; immunostimulant; immunomodulator; antiinflammatory; immunosuppressive; nephrotropic; antigout; thyromimetic; cytostatic; antibacterial; fungicide; protozoacide; antiatherosclerotic; antipolynucleotide and virucide; antipsoriatic; and hepatotropic. PPIM polynucleotide and protein sequences can be used in the diagnosis, treatment and prevention of autoimmune/inflammatory disorders such as AIDS, DiGeorge's syndrome, severe combined immunodeficiency disease (SCID), Chediak-Higashi syndrome, Cushing's disease, Addison's disease, autoimmune thyroiditis, Crohn's disease, diabetes mellitus, Good pasture's syndrome, gout, Grave's disease, Hashimoto's thyroiditis, Sjogren's syndrome, Werner's syndrome, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections and cell proliferative disorder such as arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, psoriasis and cancer. PPIM polynucleotide sequences can be used in somatic or germline gene therapy and in diagnosis of diseases. They can also be used in generating genomic hybridisation probes useful in mapping the naturally occurring genomic sequences and in molecular biology techniques

XX Sequence 775 AA;

Query Match 100.0%; Score 4070; DB 4; Length 775;
Best Local Similarity 100.0%; Pred. No. 1.8e-310;
Matches 775; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MTIVDKASESDPSAYNQPGSSEAVSPGDMAGSAGVSSINDVSNHTLSLGPVPGA 60
1 MTIVDKASESDPSAYNQPGSSEAVSPGDMAGSAGVSSINDVSNHTLSLGPVPGA 60
61 VVYSSSVDPKSPKSPKQDQALGDIAPPOKVPFSEKI CLKWQTHRVGAGLQNLGNTC 120
61 VVYSSSVDPKSPKSPKQDQALGDIAPPOKVPFSEKI CLKWQTHRVGAGLQNLGNTC 120
121 FANAALQCLTYTPPLANYMLSHESKTCBAEGFCMCTMQAHITQALSPGCDVTKPMFVI 180
121 FANAALQCLTYTPPLANYMLSHESKTCBAEGFCMCTMQAHITQALSPGCDVTKPMFVI 180
181 NEMRIARHLRFGNOEADHEFLQTVTDAMQACLSNGSKLDRHTQATTLVCOIFGGYLRS 240
181 NEMRIARHLRFGNOEADHEFLQTVTDAMQACLSNGSKLDRHTQATTLVCOIFGGYLRS 240
241 RVKCLNCKGVSDTFDYLDTLEIKAAQSVNKALEQFVKPQLDGENSYKCSKCKMYPVA 300
241 RVKCLNCKGVSDTFDYLDTLEIKAAQSVNKALEQFVKPQLDGENSYKCSKCKMYPVA 300
301 SKRFTIHRSSNVLTLSLKRANFTGKIAKDVKYPEYLDIRPMSQPNGBPIVYLYAVL 360
301 SKRFTIHRSSNVLTLSLKRANFTGKIAKDVKYPEYLDIRPMSQPNGBPIVYLYAVL 360
361 VHTGNCAGHYFCVIKASNGLWYQNDSTVSTDSIRSVLSQAAVLFYIRSHDVNGGE 420
361 VHTGNCAGHYFCVIKASNGLWYQNDSTVSTDSIRSVLSQAAVLFYIRSHDVNGGE 420
421 LTHPTHSQSSPRPVISQVVTNKAAPGFIGPOLPSHMIKNPPLHNGTGPKLDTSPSS 480
421 LTHPTHSQSSPRPVISQVVTNKAAPGFIGPOLPSHMIKNPPLHNGTGPKLDTSPSS 480
481 MSSPNGNSVNRASPNVASVONMSVNRSSVPEHPKKQKITISHNKLPRVCOQSPN 540
481 MSSPNGNSVNRASPNVASVONMSVNRSSVPEHPKKQKITISHNKLPRVCOQSPN 540
541 LHSNLENTPKVPSTTNSAVQSTNSASTMSVSSKVTKPIPRSSCSQPVNMGSKLN 600
541 LHSNLENTPKVPSTTNSAVQSTNSASTMSVSSKVTKPIPRSSCSQPVNMGSKLN 600
601 SSVLVPGAESSDSDESKGLGKENGIGTIVSSHSPGQDAEDATPHELPQMTLNGA 660
601 SSVLVPGAESSDSDESKGLGKENGIGTIVSSHSPGQDAEDATPHELPQMTLNGA 660

661 NSADSDSPKENGAPDAGASCOGQPALHSENPFAKANGLPGLMPAPLLSLPEDKILETF 720
661 NSADSDSPKENGAPDAGASCOGQPALHSENPFAKANGLPGLMPAPLLSLPEDKILETF 720
721 RLSNKLKSTDEMSAPGAERGPEDRDAEPDPOGSPAESLEPDDAAASLFFPSEFSEG 775
721 RLSNKLKSTDEMSAPGAERGPEDRDAEPDPOGSPAESLEPDDAAASLFFPSEFSEG 775
RESULT 2
AAU98889
ID AAU98889 standard; proteun; 1251 AA.
XX AAU98889;
XX
DT 26-AUG-2002 (first entry)
XX
DE Human protease PR75.
XX
XX Protease; human; gastrointestinal disorder; gastritis; atherosclerosis;
XX ulcerative colitis; Reye's syndrome; cardiovascular disease; enzyme;
XX hypertension; myocardial infarction; autoimmune disease; AIDS; PR75;
XX inflammatory disorder; acquired immunodeficiency syndrome; asthma;
XX Grave's disease; cell proliferative disorder; hepatitis; psoriasis;
XX leukaemia; developmental disorder; Cushing's syndrome; impotence;
XX epithelial disorder; dermatitis; scabies; eczema; neurological disorder;
XX Parkinson's disease; dementia; Alzheimer's disease; infertility;
XX Huntington's disease; multiple sclerosis; reproductive disorder.
XX
XX Homo sapiens.
XX
XX WO200238744-A2.
XX
XX 16-MAY-2002.
XX
XX 18-OCT-2001; 2001WO-US051034.
XX
XX 18-OCT-2000; 2000US-0241573P.
XX 25-OCT-2000; 2000US-0243643P.
XX 02-NOV-2000; 2000US-0245256P.
XX 13-NOV-2000; 2000US-0246395P.
XX 16-NOV-2000; 2000US-0249826P.
XX 20-NOV-2000; 2000US-0252303P.
XX 01-DEC-2000; 2000US-0250981P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Lee EA, Hafalia AJA, Yue H, Lal PG, Yao MG, Lu Y, Wallia NK;
XX Warren BA, Lu DAM, Baughn MR, Delegeane AM, Burford N, Borowsky ML;
XX Lee S, Xu Y, Griffin JA, Kalliock DA, Gandhi AR, Arvizu C, Ison CH;
XX Tang Y, Azimzai Y, Elliott VS, Swarnakar A, Ramkumar J, Nguyen DB;
XX Tribouley CM, Lo TP, Au-Young J, Thangavelu K, Kearney L;
XX
XX WPI; 2002-463471/49.
XX N-PSDB; ABK86140.
XX
XX New human proteases useful for diagnosing, preventing or treating
XX anorexia, myocardial infarction, Addison's disease, hepatitis, Cushing's
XX syndrome, eczema, Parkinson's disease, and impotence.
XX
XX Claim 1; Page 140-143; 168pp; English.
XX
XX This invention relates to the DNA and protein sequences of novel isolated
XX human proteases (PR75). The protein sequences of the invention are useful
XX for screening a compound for effectiveness as an agonist or antagonist of
XX its activity, the identified agonist and antagonist are useful for
XX treating a disease or condition associated with decreased or
XX overexpression of functional PR75 in a patient. The PR75 protein is also
XX useful as an immunogen for preparing polyclonal or monoclonal antibodies
XX by hybridoma technology. An antibody that binds the PR75 proteins is
XX useful for detection and purification of the proteins and can be used to
XX diagnose a condition or disease associated with expression of PR75 in a
XX subject or in a biological sample. The sequences of the invention are

CC useful for diagnosis, treatment and prevention of gastrointestinal
 CC disorders such as gastritis, ulcerative colitis, Reye's syndrome, etc;
 CC cardiovascular such as atherosclerosis, hypertension, myocardial
 CC infarction, etc; autoimmune/inflammatory disorders such as acquired
 CC immunodeficiency syndrome (AIDS), asthma, Grave's disease, etc; cell
 CC proliferative disorders such as hepatitis, psoriasis, leukaemia etc;
 CC developmental disorders such as Cushing's syndrome; epithelial disorders
 CC such as dermatitis, scabies, eczema, etc; neurological disorders such as
 CC Parkinson's disease, dementia, Alzheimer's disease, Huntington's disease,
 CC multiple sclerosis, etc; or reproductive disorders such as infertility,
 CC impotence. The present sequence represents the PRS7 protein sequence of
 CC the invention
 CC
 XX

SQ Sequence 1251 AA;

Query Match 99.0%; Score 4028; DB 5; Length 1251;
 Best Local Similarity 99.7%; Pred. No. 7.4e-307;
 Matches 767; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTIVDKASESSDPSSAYQNPQGSSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVPGA 60
 DB |||||
 QY 1 MTIVDKASESSDPSSAYQNPQGSSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVPGA 60
 DB |||||
 QY 61 VYSSSSVDPKSPQKQDQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
 DB |||||
 QY 61 VYSSSSVDPKSPQKQDQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
 DB |||||
 QY 121 FANAALQCLTYTPPLANYMLSHSHTKCHAEFCMCTQOAHITQALSNPQGVKPMFVI 180
 DB |||||
 QY 121 FANAALQCLTYTPPLANYMLSHSHTKCHAEFCMCTQOAHITQALSNPQGVKPMFVI 180
 DB |||||
 QY 181 NEMRIARHLRFNGBDAHEFLQYTVDMQKACLNGSNKLDRTHTQATTLVCOIFGGYLS 240
 DB |||||
 QY 181 NEMRIARHLRFNGBDAHEFLQYTVDMQKACLNGSNKLDRTHTQATTLVCOIFGGYLS 240
 DB |||||
 QY 241 RVKCLNCKGVSDFDPYDLITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSKCKKWPA 300
 DB |||||
 QY 241 RVKCLNCKGVSDFDPYDLITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSKCKKWPA 300
 DB |||||
 QY 301 SKRFTTHRSSNVLTLILKRFANFTGKIAKVPEYDLIRPMSQNPGEIVVLYAVL 360
 DB |||||
 QY 301 SKRFTTHRSSNVLTLILKRFANFTGKIAKVPEYDLIRPMSQNPGEIVVLYAVL 360
 DB |||||
 QY 361 VHTGFNCHAGHYFCYIKASNGLWQNDISVTSDIRSVLSQAVVLFYIRSHDVKNGE 420
 DB |||||
 QY 361 VHTGFNCHAGHYFCYIKASNGLWQNDISVTSDIRSVLSQAVVLFYIRSHDVKNGE 420
 DB |||||
 QY 421 LTHPTHSPQSSPRPVISQRVVTVNKAAPGFIGPQLPSHMIKNPPLHNGTGLKDTPSS 480
 DB |||||
 QY 421 LTHPTHSPQSSPRPVISQRVVTVNKAAPGFIGPQLPSHMIKNPPLHNGTGLKDTPSS 480
 DB |||||
 QY 481 MSSPNGNSVNRASPVNASVQNSVNRSSVPEHPKQKTIISHNKLPVRQCSQPN 540
 DB |||||
 QY 481 MSSPNGNSVNRASPVNASVQNSVNRSSVPEHPKQKTIISHNKLPVRQCSQPN 540
 DB |||||
 QY 541 LHSNSLENPTKVPVPSSTITNSAVQTSNASTMSVSKVTKPIPRSESCSQPMNCKSKLN 600
 DB |||||
 QY 541 LHSNSLENPTKVPVPSSTITNSAVQTSNASTMSVSKVTKPIPRSESCSQPMNCKSKLN 600
 DB |||||
 QY 601 SSVLPYGAESSEDDESKGLGKENGIGTIVSSHSPQDADEEATPHELQEPMTLNGA 660
 DB |||||
 QY 601 SSVLPYGAESSEDDESKGLGKENGIGTIVSSHSPQDADEEATPHELQEPMTLNGA 660
 DB |||||
 QY 661 NSADSDPDKENGLAPDGCQGPALHSENPFAKANGLPGLKMPAPLLSLPEDKILTF 720
 DB |||||
 QY 661 NSADSDPDKENGLAPDGCQGPALHSENPFAKANGLPGLKMPAPLLSLPEDKILTF 720
 DB |||||
 QY 721 RLSNKLKGTSDMSAPGAERGPPEDRDPAEPQPGSPAESLEEDPDAASL 769
 DB |||||
 QY 721 RLSNKLKGTSDMSAPGAERGPPEDRDPAEPQPGSPAESLEEDPDAASL 769
 DB |||||

RESULT 3

AAB95146
 ID AAB95146 standard; protein; 1197 AA.
 XX
 AC AAB95146;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:17169.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EF1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-00116126.
 XX
 PR 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
 PT
 PS Claim 8; SEQ ID NO 17169; 2537pp + Sequence Listing; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
 CC
 CC Sequence 1197 AA;

Query Match 98.8%; Score 4022; DB 4; Length 1197;
 Best Local Similarity 99.6%; Pred. No. 2e-306;
 Matches 766; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MTIVDKASESSDPSSAYQNPQGSSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVPGA 60
 DB |||||
 QY 1 MTIVDKASESSDPSSAYQNPQGSSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVPGA 60
 DB |||||
 QY 61 VYSSSSVDPKSPQKQDQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
 DB |||||

Db	61	VVYSSSSVDPKSKPSQKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVAGLQNLGNTC	120
Qy	121	FANAALQCLTYTPPLANYMLSHSHSKTCHAEFGCMCTMQAHITQALSNGPDVVKPMFVI	180
Db	121	FANAALQCLTYTPPLANYMLSHSHSKTCHAEFGCMCTMQAHITQALSNGPDVVKPMFVI	180
Qy	181	NEMRRIARHLRFQNGQDAHEFLQYTVDMQKACLNGSKNLDHRTQATTLVCQIFGGYLRS	240
Db	181	NEMRRIARHLRFQNGQDAHEFLQYTVDMQKACLNGSKNLDHRTQATTLVCQIFGGYLRS	240
Qy	241	RVKCLNCKGVSDTFDYPYLDITLTKAASVKNKALEQFVKPEQLDGENSYKSKCKKWVPA	300
Db	241	RVKCLNCKGVSDTFDYPYLDITLTKAASVKNKALEQFVKPEQLDGENSYKSKCKKWVPA	300
Qy	301	SKRFTIHRSSNVLTLSLKRANFTGGKIAKDVKYPEYLDIRPYMSQPNGEPIVVLYAVL	360
Db	301	SKRFTIHRSSNVLTLSLKRANFTGGKIAKDVKYPEYLDIRPYMSQPNGEPIVVLYAVL	360
Qy	361	VHTGFNCHAGHYFCYIKASNGLWYQMNDSIVSTDIRSVLSQAYVLFYIRSHDVKNNGE	420
Db	361	VHTGFNCHAGHYFCYIKASNGLWYQMNDSIVSTDIRSVLSQAYVLFYIRSHDVKNNGE	420
Qy	421	LTHPTHSPGSSRPVTSQRVWTKQAAPGFIGPQLPSHMIKNPPLHNGTGPKLDTSPSS	480
Db	421	LTHPTHSPGSSRPVTSQRVWTKQAAPGFIGPQLPSHMIKNPPLHNGTGPKLDTSPSS	480
Qy	481	MSSPNGNSSVNRASPVNASQVWNSVNRSSVPEHPKPKKITISHNKLPVRQCQOPN	540
Db	481	MSSPNGNSSVNRASPVNASQVWNSVNRSSVPEHPKPKKITISHNKLPVRQCQOPN	540
Qy	541	LHNSLENTPKVPSSITTSNAVOSTSNASTMSVSKVTKEIPRESESCOPVWNGKSKLN	600
Db	541	LHNSLENTPKVPSSITTSNAVOSTSNASTMSVSKVTKEIPRESESCOPVWNGKSKLN	600
Qy	601	SSVLVPYGASSEDSDSESKGLKENGIGTIVSHSPQDAEDDEATPHEIQEPTLNGA	660
Db	601	SSVLVPYGASSEDSDSESKGLKENGIGTIVSHSPQDAEDDEATPHEIQEPTLNGA	660
Qy	661	NSADSDDPKENGAPDGLAPDGCQOPALHSENPFAKANGLPKMLPAPLLSLPEDKILETF	720
Db	661	NSADSDDPKENGAPDGLAPDGCQOPALHSENPFAKANGLPKMLPAPLLSLPEDKILETF	720
Qy	721	RLSNKLGKSTDEMSAPGAERGPEDRDAPQPGSPAESLEEDPAAASL	769
Db	721	RLSNKLGKSTDEMSAPGAERGPEDRDAPQPGSPAESLEEDPAAASL	769
RESULT 4			
Id	ABP69291	standard; protein; 1270 AA.	
XX	AC	ABP69291;	
XX	AC	ABP69291;	
DT	20-JAN-2003	(first entry)	
DE	XX	Human polypeptide SEQ ID NO 1338.	
XX	KW	Human; genome mapping; gene therapy; food supplement; virus; fungus;	
XX	KW	cell-proliferative disorder; neurodegenerative disease; bacterial;	
XX	KW	Parkinson's disease; Alzheimer's disease; autoimmune disease;	
XX	KW	multiple sclerosis; diabetes; genetic disorder; wound; infection;	
XX	KW	arthritis; cytosolic; immunomodulator; neurotropic; neuroprotective;	
XX	KW	antiparkinsonian; antidiabetic; immunosuppressive; dermatological;	
XX	KW	haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;	
XX	XX	antiarthritic.	
OS	XX	Homo sapiens.	
PN	XX	WO200270539-A2.	
PD	XX	12-SEP-2002.	
XX	XX		

PF	05-MAR-2002;	2002W0-US003095.	
XX	05-MAR-2001;	2001US-00799451.	
XX	(HYSE-) HYSEQ INC.		
XX	Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;		
PI	Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;		
PI	Wehrman T, Wang J, Wang D, Drmanac RT;		
XX	WPI; 2002-759812/82.		
DR	N-PSDB; AB211508.		
XX	New polynucleotides comprising sequences assembled from expressed		
PT	sequence tags (ESTs). useful for treating cell-proliferative,		
PT	neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet		
PT	or coagulation disorders.		
XX	Claim 9; SEQ ID NO 1338; 1012pp + Sequence Listing; English.		
XX	The invention relates to an isolated polynucleotide (I) comprising a		
CC	nucleotide sequence selected from any of 948 sequences (AB211119-		
CC	AB212066) or their mature protein coding portion, active domain coding		
CC	protein or complementary sequences. The polynucleotides are useful for		
CC	identifying expressed genes or for physical mapping of human genome. The		
CC	encoded polypeptides (AEP6902-ABP6949) are useful as molecular weight		
CC	markers, as a food supplement, for generating antibodies, in medical		
CC	imaging, screening and diagnostic assays and for treating cell-		
CC	proliferative disorders (cancer), neurodegenerative diseases (Parkinson's		
CC	or Alzheimer's disease), autoimmune diseases (multiple sclerosis,		
CC	diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,		
CC	platelet or coagulation disorders, wound, burns, incision, ulcers, liver		
CC	or lung fibrosis, infections (bacterial, viral, fungal, parasitic),		
CC	arthritis, etc. Note: The sequence data for this patent did not form part		
CC	of the printed specification, but was obtained in electronic format		
CC	directly from WIPO at ftp.wipo.int/pub/published_pct_sequences		
XX	Sequence 1270 AA;		
Qy	Query Match	98.8%; Score 4021; DB 5; Length 1270;	
Db	Best Local Similarity	99.6%; Pred. No. 2.7e-306;	
Qy	Matches	766; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
Db	1	MTIYDKASESDPSAYQNPQGSSEAVSPGDMAGSAGVSSLDVSNHTLSLGPVPGA	60
Qy	61	VVYSSSSVDPKSKPSQKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVAGLQNLGNTC	120
Db	61	VVYSSSSVDPKSKPSQKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVAGLQNLGNTC	120
Qy	121	FANAALQCLTYTPPLANYMLSHSHSKTCHAEFGCMCTMQAHITQALSNGPDVVKPMFVI	180
Db	121	FANAALQCLTYTPPLANYMLSHSHSKTCHAEFGCMCTMQAHITQALSNGPDVVKPMFVI	180
Qy	181	NEMRRIARHLRFQNGQDAHEFLQYTVDMQKACLNGSKNLDHRTQATTLVCQIFGGYLRS	240
Db	181	NEMRRIARHLRFQNGQDAHEFLQYTVDMQKACLNGSKNLDHRTQATTLVCQIFGGYLRS	240
Qy	241	RVKCLNCKGVSDTFDYPYLDITLTKAASVKNKALEQFVKPEQLDGENSYKSKCKKWVPA	300
Db	241	RVKCLNCKGVSDTFDYPYLDITLTKAASVKNKALEQFVKPEQLDGENSYKSKCKKWVPA	300
Qy	301	SKRFTIHRSSNVLTLSLKRANFTGGKIAKDVKYPEYLDIRPYMSQPNGEPIVVLYAVL	360
Db	301	SKRFTIHRSSNVLTLSLKRANFTGGKIAKDVKYPEYLDIRPYMSQPNGEPIVVLYAVL	360
Qy	361	VHTGFNCHAGHYFCYIKASNGLWYQMNDSIVSTDIRSVLSQAYVLFYIRSHDVKNNGE	420
Db	361	VHTGFNCHAGHYFCYIKASNGLWYQMNDSIVSTDIRSVLSQAYVLFYIRSHDVKNNGE	420
Qy	421	LTHPTHSPGSSRPVTSQRVWTKQAAPGFIGPQLPSHMIKNPPLHNGTGPKLDTSPSS	480

Db 421 LTHPTSPGQSPRPVISQRVVTKQAAPGIGFQPLPSHMKNPPLHNGTGLKDTTPSS 480
QY 481 MSSPNGSSVNRASPVNASVQNSVNRSSVPEHPKKQKITIISHNKLPVRQCSQPN 540
Db 481 MSSPNGSSVNRASPVNASVQNSVNRSSVPEHPKKQKITIISHNKLPVRQCSQPN 540
QY 541 LHSNSLENPTKPVPSSTITNSAVQSTNSASTMSVSSVTKPIPRSESCSPVNMGSKLN 600
Db 541 LHSNSLENPTKPVPSSTITNSAVQSTNSASTMSVSSVTKPIPRSESCSPVNMGSKLN 600
QY 601 SSVLVPGAESSESDSEESKGLKENGIGTIVSSHSPGQDADEATPHLOPMTINGA 660
Db 601 SSVLVPGAESSESDSEESKGLKENGIGTIVSSHSPGQDADEATPHLOPMTINGA 660
QY 661 NSADSDSDPKENGLAPDGASCOQOPALHSENPFKANGLPGLKMPAPLLSLPDKILETF 720
Db 661 NSADSDSDPKENGLAPDGASCOQOPALHSENPFKANGLPGLKMPAPLLSLPDKILETF 720
QY 721 RLSNKLKSGTDEMSAPGAERGPPEDRDAEPQGPSAAESLEEPPDAASL 769
Db 721 RLSNKLKSGTDEMSAPGAERGPPEDRDAEPQGPSAAESLEEPPDAAGL 769
RESULT 5
AAB82177
ID AAB82177 standard; protein; 762 AA.
AC AAB82177;
XX
DT 20-JUL-2001 (first entry)
XX
DE Human ubiquitin protease 23431.
KW Human; ubiquitin protease; deubiquitinating enzyme; tuberculostatic;
KW antiasthmatic; antiinflammatory; antidiarrhoeic; hepatotropic;
KW gynaecological; cytostatic; antimicrobial; neuroprotective; anti-HIV;
KW immunosuppressive; cardiant; antianaemic; nephrotropic; antibacterial;
KW anti-thyroid; gastrointestinal.
XX
OS Homo sapiens.
XX
FN WO200123589-A2.
XX
PD 05-APR-2001.
XX
PF 29-SEP-2000; 2000WO-US026962.
XX
PR 29-SEP-1999; 99US-00407356.
XX
XX (MILL-) MILLENNIUM PHARM INC.
FA Kapeller-Libermann R;
XX
PI
PI
DR WPI; 2001-374253/39.
DR N-PSDB; AAH19322.
XX
PT New human ubiquitin protease, a member of the mammalian deubiquitinating
PT enzymes is useful for diagnosis and treatment of e.g. tuberculosis and
PT Alzheimer's disease.
XX
XX Claim 5; Fig 1; 111pp; English.
XX
XX The present sequence is human ubiquitin protease 23431. The ubiquitin
XX protease coding sequence and protein are useful as reagents or targets in
XX assays for treatment and diagnosis of ubiquitin-mediated or -related
XX disorders, especially disorders mediated by deubiquitinating enzymes. The
XX protein and coding sequence are also useful for treating disorders
XX involving the following: the spleen e.g. tuberculosis and typhoid fever,
XX the lung such as bronchial asthma, the colon e.g. as diarrhoea and
XX dysentery, the liver e.g. jaundice and cholestasis, the uterus and
XX endometrium e.g. endometriosis, the brain e.g. acute meningitis and
XX Alzheimer disease, T-cells including transplant rejection and autoimmune
XX diseases such as systemic lupus erythematosus, diseases of the skin such

CC as malignant melanoma, the bone marrow e.g. B- and T-lymphoid leukaemias,
CC the heart including myocardial infarction, red cells e.g. anaemia, the
CC thymus e.g. Hodgkin disease, B-cells e.g. peripheral B-cell neoplasms,
CC the kidney e.g. polycystic kidney disease, the breast including
CC periductal mastitis, the testis and epididymis e.g. syphilis, the
CC prostate e.g. nodular hyperplasia, the thyroid, e.g. hyperthyroidism, the
CC skeletal muscle e.g. rhabdomyosarcoma, the pancreas e.g. ectopic
CC pancreas, reduced platelet number e.g. HIV (human immunodeficiency virus)
CC -associated thrombocytopenia and disorders involving precursor T-cell
CC neoplasms including precursor T lymphoblastic leukaemia/lymphoma
XX
SQ Sequence 762 AA;
Query Match 98.3%; Score 3999; DB 4; Length 762;
Best Local Similarity 99.9%; Pred. No. 6.e-305;
Matches 761; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MTIVDKASSESDPSAYONQPGSEAVSPGDMAGSASWGAVSLNDVSNHTLSLGPVPGA 60
Db 1 MTIVDKASSESDPSAYONQPGSEAVSPGDMAGSASWGAVSLNDVSNHTLSLGPVPGA 60
QY 61 VVYSSSVDPDKSPQKQDQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
Db 61 VVYSSSVDPDKSPQKQDQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTTPPLANYMLSHESKTCBAEGFCWMCCTMQAHTQALSNPQGVKPMFVI 180
Db 121 FANAALQCLTYTTPPLANYMLSHESKTCBAEGFCWMCCTMQAHTQALSNPQGVKPMFVI 180
QY 181 NEMRRIARHLRFNGQEDAHEFLQYTVDMQKACLNGSNKLDHRHTQATTLVQCFGGYLSR 240
Db 181 NEMRRIARHLRFNGQEDAHEFLQYTVDMQKACLNGSNKLDHRHTQATTLVQCFGGYLSR 240
QY 241 RVKCLNCKGVSDFPDYDITLLEIKAAQSVNKALQFVKPEQLDGENSKYCKCKKWVPA 300
Db 241 RVKCLNCKGVSDFPDYDITLLEIKAAQSVNKALQFVKPEQLDGENSKYCKCKKWVPA 300
QY 301 SKRFTIHRSSNVLTLSLKRFAFTGCKIAKDVKPYEYLDIRPYMSQNPGEPIVYVYAVL 360
Db 301 SKRFTIHRSSNVLTLSLKRFAFTGCKIAKDVKPYEYLDIRPYMSQNPGEPIVYVYAVL 360
QY 361 VHTGENCHAGHYFCYIKASNGLWYQNDISIVSTDIRSVLSQOAYVLYFIRSHDVKNGE 420
Db 361 VHTGENCHAGHYFCYIKASNGLWYQNDISIVSTDIRSVLSQOAYVLYFIRSHDVKNGE 420
QY 421 LTHPTSPGQSPRPVISQRVVTKQAAPGIGFQPLPSHMKNPPLHNGTGLKDTTPSS 480
Db 421 LTHPTSPGQSPRPVISQRVVTKQAAPGIGFQPLPSHMKNPPLHNGTGLKDTTPSS 480
QY 481 MSSPNGSSVNRASPVNASVQNSVNRSSVPEHPKKQKITIISHNKLPVRQCSQPN 540
Db 481 MSSPNGSSVNRASPVNASVQNSVNRSSVPEHPKKQKITIISHNKLPVRQCSQPN 540
QY 541 LHSNSLENPTKPVPSSTITNSAVQSTNSASTMSVSSVTKPIPRSESCSPVNMGSKLN 600
Db 541 LHSNSLENPTKPVPSSTITNSAVQSTNSASTMSVSSVTKPIPRSESCSPVNMGSKLN 600
QY 601 SSVLVPGAESSESDSEESKGLKENGIGTIVSSHSPGQDADEATPHLOPMTINGA 660
Db 601 SSVLVPGAESSESDSEESKGLKENGIGTIVSSHSPGQDADEATPHLOPMTINGA 660
QY 661 NSADSDSDPKENGLAPDGASCOQOPALHSENPFKANGLPGLKMPAPLLSLPDKILETF 720
Db 661 NSADSDSDPKENGLAPDGASCOQOPALHSENPFKANGLPGLKMPAPLLSLPDKILETF 720
QY 721 RLSNKLKSGTDEMSAPGAERGPPEDRDAEPQGPSAAESLEE 762
Db 721 RLSNKLKSGTDEMSAPGAERGPPEDRDAEPQGPSAAESLEE 762
RESULT 6
ABU08951
ID ABU08951 standard; protein; 762 AA.

CC	(e.g. ischaemic heart disease, myocardial infarction and cardiomyopathy),
CC	blood vessel disorders (e.g. atherosclerosis, thrombophlebitis and
CC	Raynaud disease), red blood cell disorders (e.g. anaemia), thymus
CC	disorders (e.g. DiGeorge syndrome), B-cell disorders (e.g. leukaemia),
CC	kidney disorders (e.g. polycystic kidney disease and glomerulonephritis),
CC	breast disorders (e.g. mastitis), testicular disorders (e.g. sexually
CC	transmitted diseases and cryptorchidism), thyroid disorders (e.g.
CC	hypothyroidism), pancreatic disorders (e.g. pancreatitis), and intestinal
CC	disorders (e.g. whipple disease), as well as tumours and cancers of the
CC	above listed organs/cells. Many more diseases and disorders are listed in
CC	the specification. The present sequence represents human ubiquitin
XX	protease
XX	Sequence 762 AA;
XX	Query Match 98.3%; Score 3999; DB 6; Length 762;
XX	Best Local Similarity 99.9%; Pred. No. 6.6e-305; Indels 0; Gaps 0;
XX	Matches 761; Conservative 0; Mismatches 1;
QY	1 MTIVDKASESSDPSAYQNQPGSSEAVSPGMDAGSASGAVSSLNDVSNHTLSLGPVPGA 60
DB	1 MTIVDKASESSDPSAYQNQPGSSEAVSPGMDAGSASGAVSSLNDVSNHTLSLGPVPGA 60
QY	61 VYSSSSVDPKSPQKQDQALGDGIAPQKVLFPSEKICLKWOOTHRVCGALQNGTC 120
DB	61 VYSSSSVDPKSPQKQDQALGDGIAPQKVLFPSEKICLKWOOTHRVCGALQNGTC 120
QY	121 FANAALQCLTYTTPPLANYMLSHESKTCIAEGFQWCTMOAHITQALSNFGDVIKPMFVI 180
DB	121 FANAALQCLTYTTPPLANYMLSHESKTCIAEGFQWCTMOAHITQALSNFGDVIKPMFVI 180
QY	181 NEMRRIARHLRFQNGEDAHEFLQYTVDMQKACLNGSKLDRHTQATTLVCOIFGGVYLS 240
DB	181 NEMRRIARHLRFQNGEDAHEFLQYTVDMQKACLNGSKLDRHTQATTLVCOIFGGVYLS 240
QY	241 RVKCLNCKGVSDTFTPYLDITLBIKAAQSVNKALEQFVKPEQLDGENSYKSCCKKWMPA 300
DB	241 RVKCLNCKGVSDTFTPYLDITLBIKAAQSVNKALEQFVKPEQLDGENSYKSCCKKWMPA 300
QY	301 SKRFTHRSSNVLTLILKRFANFTGGKIADVKYPEYLDIRPYMSQNGEPIVTVLYAVL 360
DB	301 SKRFTHRSSNVLTLILKRFANFTGGKIADVKYPEYLDIRPYMSQNGEPIVTVLYAVL 360
QY	361 VHTGFNCHAGHYFCYIKASNGLWYQNDISIVTSDIRSVLSQQAYVLYFIRSHDVKNGE 420
DB	361 VHTGFNCHAGHYFCYIKASNGLWYQNDISIVTSDIRSVLSQQAYVLYFIRSHDVKNGE 420
QY	421 LTHPTSPGQSSPRPVIQSQRVVTNKAAPGFIGPOLPSHMIKNPPLNGTGLKDTPTSS 480
DB	421 LTHPTSPGQSSPRPVIQSQRVVTNKAAPGFIGPOLPSHMIKNPPLNGTGLKDTPTSS 480
QY	481 MSSPENGSSVNRASPVNASVQWNSVNRSSVIEHPKPKQKITISIHNLKLPVRCQSQPN 540
DB	481 MSSPENGSSVNRASPVNASVQWNSVNRSSVIEHPKPKQKITISIHNLKLPVRCQSQPN 540
QY	541 LHSNLENPTKVPVPSITITNSAVQSTSNASTMSVSSKVTXPIPRSESCSQPMWNGSKLN 600
DB	541 LHSNLENPTKVPVPSITITNSAVQSTSNASTMSVSSKVTXPIPRSESCSQPMWNGSKLN 600
QY	601 SSVLPVYGAESSESDDESKGLXENGIGTIVSSHSPGQDAEDEATPHEIQEPMTLNGA 660
DB	601 SSVLPVYGAESSESDDESKGLXENGIGTIVSSHSPGQDAEDEATPHEIQEPMTLNGA 660
QY	661 NSADSDDPKENGLAPDQASQGOQALHSENPFKANGLPGLKMPAPLLSPEDKILETF 720
DB	661 NSADSDDPKENGLAPDQASQGOQALHSENPFKANGLPGLKMPAPLLSPEDKILETF 720
QY	721 RLSNKLKSTDEMSAPGAERGPEDRDAPPGSPAAESLEE 762
DB	721 RLSNKLKSTDEMSAPGAERGPEDRDAPPGSPAAESLEE 762
XX	RESULT 7

XX Human; enzyme; ubiquitin protease; spleen disorder; Hodgkin's disease;
AC lung disorder; adult respiratory distress syndrome; colon disorder;
XX inflammatory bowel disease; liver disorder; jaundice; uterine disorder;
DT endometriosis; brain disorder; Alzheimer's disease; T-cell disorder;
XX acquired immunodeficiency syndrome; AIDS; skin disorder; urticaria;
DE heart disorder; ischaemic heart disease; blood vessel disorder;
XX atherosclerosis; red blood cell disorder; anaemia; thymus disorder;
KW DiGeorge syndrome; B-cell disorder; leukaemia; kidney disorder;
KW polycystic kidney disease; glomerulonephritis; breast disorder; mastitis;
KW testicular disorder; sexually transmitted disease; thyroid disorder;
KW hypothyroidism; pancreatic disorder; pancreatitis; intestinal disorder;
KW whipple disease; tumour; cancer.
XX Homo sapiens.
XX US2003037350-A1.
XX 20-FEB-2003.
XX 05-JUN-2002; 2002US-00163547.
XX 29-SEP-1999; 99US-00407356.
XX 03-NOV-1999; 99US-00435311.
XX 01-FEB-2000; 2000US-00496005.
XX 11-FEB-2000; 2000US-01824009P.
XX 14-FEB-2000; 2000US-0182408P.
XX 28-FEB-2000; 2000US-0185503P.
XX 02-NOV-2000; 2000US-00704918.
XX 12-FEB-2001; 2001US-00781598.
XX 14-FEB-2001; 2001US-00782952.
XX 28-FEB-2001; 2001US-00796100.
XX (MILL-) MILLENNIUM PHARM INC.
XX Glucksmann MA, Kapeller-Libermann R, Meyers RE, Rudolph-Owen LA;
XX WPI; 2003-342047/32.
XX N-PSDB; ABX93880.
XX New nucleic acid or polypeptide, useful for preparing a composition for
XX diagnosing or treating diseases e.g., tumor.
XX Claim 9; Fig 1; 281pp; English.
XX The invention relates to a new isolated nucleic acid molecule encoding
XX one of 6 polypeptides (ubiquitin protease, lipase, Dynamain, short chain
XX dehydrogenase, ADAM-TS (a disintegrin and metalloprotease domain protein
XX with thrombospondin (TS) domains) and gamma butyrobetaine-hydroxylase
XX (gamma BBH). Also included are the polypeptide, host cells containing the
XX nucleic acids, an antibody that selectively binds to the polypeptide, a
XX method for producing the polypeptide, a method for detecting the presence
XX of the polypeptide or the nucleic acid in a sample, a method for
XX identifying a compound that binds to the polypeptide, a method for
XX modulating the activity of the polypeptide and a method for identifying a
XX compound that modulates the activity of the polypeptide. The nucleic acid
XX or polypeptide is useful for preparing a composition for diagnosing or
XX treating diseases e.g. spleen disorders (e.g. splenomegaly and Hodgkin's
XX disease), lung disorders (e.g. adult respiratory distress syndrome,
XX pulmonary oedema, chronic bronchitis and emphysema), colon disorders
XX (e.g. stenosis, colitis, inflammatory bowel disease and Crohn's disease),
XX liver disorders (e.g. jaundice, cirrhosis, hepatitis and alcoholic liver
XX disease), uterine and endometrial disorders (e.g. endometriosis and
XX menopausal changes), brain disorders (e.g. encephalitis, Alzheimer's
XX disease, Parkinson's disease, ataxia and multiple sclerosis), T-cell
XX disorders (e.g. acquired immunodeficiency syndrome, AIDS), skin disorders
XX (e.g. urticaria, dermatitis and lupus erythematosus), heart disorders

XX	(INCY-) INCYTE GENOMICS INC.	
PA	Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GE, Jackson JL;	
XX	Yu JY, Tuason O, Yap PE, Amshay SR, Dam TC, Liu TF, Gerstin EH;	
PI	Peralta CH, Lewis SA, Chen A, Marwaha R, Lan RY, Urashka ME;	
PI	Kristnam SR, Kolluru V, Panesar IS;	
XX	WPI: 2003-636732/60.	
DR	N-PSDB; ADE31255.	
XX	New human diagnostic and therapeutic polynucleotides and polypeptides,	
PT	useful for diagnosing, treating or preventing e.g. leukemia, brain	
PT	cancer, atherosclerosis, AIDS, thyroiditis, infections, obesity, stroke	
PT	or Alzheimer's.	
XX	Claim 27; SEQ ID NO 198; 634pp; English.	
XX	The invention relates to a novel isolated human diagnostic and	
CC	therapeutic polynucleotide (designated dithp). The novel dithp	
CC	polynucleotide comprises: any of 188 DNA sequences consisting of 195-7798	
CC	base pairs fully defined in the specification; a polynucleotide	
CC	comprising a naturally occurring polynucleotide sequence at least 90%	
CC	identical to the dithp polynucleotide; a polynucleotide complementary to	
CC	identical; or an RNA equivalent of any of the polynucleotides mentioned	
CC	above. The dithp polynucleotides have the following activities:	
CC	antiarteriosclerotic, antiinflammatory, cerebroprotective, antilipaeamic,	
CC	antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,	
CC	tranquilizer, osteopathic, antiarthritic, antirheumatic, cytostatic,	
CC	hepatotropic, virucide, haemostatic, anti-HIV, antithyroid, thyromimetic,	
CC	dermatological, antibacterial, fungicide, antiparasitic, anticonvulsant,	
CC	thrombolytic, anticoagulant, anorectic, vasotropic, and anticulcer. The	
CC	novel DITHP polynucleotides polypeptide can be used in gene therapy and	
CC	protein replacement therapy. The dithp polynucleotides or DITHP	
CC	polypeptides are useful for diagnosing, preventing or treating diseases	
CC	associated with the expression of human molecules. In particular, these	
CC	diseases include cancers (e.g. adenocarcinoma, leukaemia, melanoma, brain	
CC	cancer, breast cancer, cervix cancer, bone cancer, liver cancer, lung	
CC	cancer) or other cell proliferative disorders (e.g. arteriosclerosis,	
CC	atherosclerosis, cirrhosis, hepatitis, polycythemia vera, primary	
CC	thrombocytopenia), autoimmune/inflammatory disorders (e.g. AIDS,	
CC	Addison's disease, thyroiditis, Crohn's disease, Graves' disease,	
CC	Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid	
CC	arthritis, scleroderma, systemic lupus erythematosus), infections (e.g.	
CC	viral, bacterial, fungal or parasitic infection), developmental disorders	
CC	(e.g. Cushing's syndrome or epilepsy), endocrine disorders (e.g.	
CC	thrombosis, hypopituitarism, hypogonadism, gigantism, goiter) metabolic	
CC	disorders (e.g. hypercholesterolemia, hypoglycaemia, diabetes,	
CC	hyperlipidaemia, obesity), neurological disorders (e.g. ischaemic	
CC	cerebrovascular disease, stroke, Alzheimer's disease, Pick's disease,	
CC	Huntington's disease, Parkinson's disease, Creutzfeldt-Jakob disease,	
CC	anxiety, schizophrenia), gastrointestinal disorders (e.g. ulcers),	
CC	transport disorders (e.g. akinesia or multidrug resistance), or	
CC	connective tissue disorders (e.g. Paget's disease or rickets). This	
XX	sequence represents one of the human DITHP polypeptides of the invention.	
XX	Sequence 1087 AA;	
XX	Query Match 35.0%; Score 1423.5; DB 7; Length 1087;	
XX	Best Local Similarity 41.2%; Pred. No. 2.7e-102;	
XX	Matches 336; Conservative 109; Mismatches 244; Indels 127; Gaps 25;	
QY	1 MTIVDKASSESDPSAYQNGSSEAVSPGDMGASGAVSSLNDV----- 47	
DB	3 MPVIVDKLEAL-----KPGKDSADDGELGKLLASSAKVLLQKIEFEPASKSFSYQL 55	
QY	48 -----SNHTLSLGPVCAVYV-SSSSVPDKSPQKDQALGDIAPPQKVLEPSEKICLK 102	
DB	56 EALKSKVLLNPKTEGASHKSGDDPPARRQSEHTYESCGDGPAPQKVLPFTKLSLR 115	
QY	103 WOOTHVRVGLNIGNTCFANALQCLTYTTPPLANTYMLSHSKTKCHARGFCNMCTMQAH 162	
DB	116 WERFVRVAGLHNLGHTCFLNATIQCLTYTTPPLANTYMLSHKSHCHQSFQMLCVMQNH 175	
QY	163 ITQALSNCEDVTKPMFVINEMERIARHLRFQNOEDAHEFLOYTVDAMQACLNGSKLDR 222	
DB	176 IVQAFANGSNAIKPVSFIRDLKIAHFRFGQDEAHEFLRYTIDAMQACLNGCAKLR 235	
QY	223 HTQATTLVCQIFGGYLRSRVKCLNCKGVSDTDPDYLDTLETKAAQSVNKALEQFVKPQ 282	
DB	236 QTAATTLVHQIFGGYLRSRVKCVSKSVSDTDPYDLIALEIRQAANI VRALELFVKADV 295	
QY	283 LDGENSKCKCKMYPASKRFTIHRSSNVLTSLKRFANFTGGKIADKVKPYEYLDIRP 342	
DB	296 LSGENAYMCAKCKKYPASKRFTIHRSSNVLTSLKRFANFTGGKIADKVKPYEYLDIRP 355	
QY	343 YMSQNGEPIVVVLYAVLHVTGCHAGHYFVYKASNLGLWQMNDISIVSTDSIRSVLSQ 402	
DB	356 YMSQNGDPMYGLYAVLVHSGYCHAGHYCYVVKASNGQWYQMNDSLHSSNVKVLNQ 415	
QY	403 QAYVLYIRSHDVYKNGELTHFTHSPG-QSSRPVTSQRVVTNKKQAAPGFIQPO-IPSHM 460	
DB	416 QAYVLYLR-----IFGSKSPREGLISR---TGSSSLPG--RESVIPDHS 455	
QY	461 IKMPPHLNGTGPL--KOTPSSSSMSSPNGNSSVNRSPVNASASVQNVSSVIRPEHPX 518	
DB	456 KKNIGNGIISPLTGKRGQSGTMKPHTEETI--GVPIKNGSTLGLKSONGCIIPKPLPS 513	
QY	519 KQKITISIHKKLPVRCQSQPNLHNSLENP-----TKPVPS-----STITNSAVOSTNAST 571	
DB	514 GS-----PSPKLSQTPTHTPILDDPGKKVKKPAPPQHPSPRTAQGLPQTSNSN- 562	
QY	572 MSVSSKVTKPIPRSESCQPVMMGSKLNSVLV-PYGAESSESDSDESKLKGKNGIGT 630	
DB	563 ---SSRSGSQRGSDRDVLTSPKLLATATANGHGLK---GNDESAGLDR-RGSSS 614	
QY	631 IVSSHSPQDAEAEATPHLQBPMTLNGANSADSDPKENGLAPDGACQOGPALHSE 690	
DB	615 SSPEHSASSDSTKAFQTPRS-----GAAHLCDSQE-----TNC--STAGHSK 654	
QY	691 NPPAKANGLPQKLM-----PAPLSLPEDKILETFRLSNKLGSTDEMSAPGAER 740	
DB	655 TPFGADSKTVKLKPVLSNTTTEPASTMPPPAK-----KLALSAKKASTLWRATG--- 706	
QY	741 GPPEDRDAEPQPGSFAAESLEEP-----DAAASLFP 771	
DB	707 -----NDLRPPPPPSVSSD-LTHPMKTHSHVVAATWP 736	
XX	RESULT 9	
XX	AAV71889	
ID	AAV71889 standard; protein; 1123 AA.	
XX	AAV71889;	
AC	13-JUN-2001 (first entry)	
DT	Human ubiquitin protease.	
DE	Human; ubiquitin protease; cytostatic; virucidal; gene therapy; tumour;	
XX	neoplasia; hepatic injury; cirrhosis; hepatitis; heart failure; rickets;	
KW	myocardial infarction; skeletal muscle tumour; Grave's disease; atrophy;	
KW	glomerulonephritis; cryotorchidism; periductal mastitis; Paget's disease;	
KW	cystic renal dysplasia; polycystic ovarian disease; cretinism; myxoedema;	
KW	inflammation; obstructive pulmonary disease; osteoporosis; enterocolitis;	
KW	idiopathic inflammatory bowel disease; bronchial carcinoma; osteomalacia;	
KW	choriocarcinoma; nodular hyperplasia; metastatic tumour; thyroiditis.	
XX	Homo sapiens.	
OS	Key	Location/Qualifiers
XX	Modified-site 13..16	/note= "Amidation site"
FT	Modified-site 15..18	/note= "cAMP and cGMP-dependent protein kinase

FT	phosphorylation site"	FT	/note= "Protein kinase C phosphorylation site"
FT	18. .21	FT	566. .568
FT	Modified-site	FT	/note= "Protein kinase C phosphorylation site"
FT	31. .33	FT	570. .575
FT	Modified-site	FT	/note= "N-myristoylation site"
FT	75. .78	FT	582. .584
FT	Modified-site	FT	/note= "Protein kinase C phosphorylation site"
FT	83. .90	FT	595. .600
FT	Modified-site	FT	/note= "N-myristoylation site"
FT	85. .90	FT	607. .610
FT	Modified-site	FT	/note= "cAMP and cGMP-dependent protein kinase phosphorylation site"
FT	92. .95	FT	609. .614
FT	Modified-site	FT	/note= "N-myristoylation site"
FT	107. .109	FT	613. .616
FT	Modified-site	FT	/note= "Casein kinase II phosphorylation site"
FT	111. .113	FT	623. .625
FT	Modified-site	FT	/note= "Protein kinase C phosphorylation site"
FT	134. .137	FT	629. .631
FT	Modified-site	FT	/note= "Protein kinase C phosphorylation site"
FT	260. .263	FT	644. .647
FT	Modified-site	FT	/note= "Asn is N-glycosylated"
FT	312. .314	FT	656. .659
FT	Modified-site	FT	/note= "Casein kinase II phosphorylation site"
FT	313. .316	FT	662. .664
FT	Modified-site	FT	/note= "Protein kinase C phosphorylation site"
FT	/note= "cAMP and cGMP-dependent protein kinase phosphorylation site"	FT	672. .675
FT	327. .329	FT	/note= "Asn is N-glycosylated"
FT	Modified-site	FT	673. .676
FT	333. .336	FT	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	FT	692. .694
FT	336. .341	FT	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	FT	694. .697
FT	338. .345	FT	/note= "cAMP and cGMP-dependent protein kinase phosphorylation site"
FT	Modified-site	FT	703. .706
FT	365. .383	FT	/note= "Casein kinase II phosphorylation site"
FT	Region	FT	748. .750
FT	signature"	FT	/note= "Protein kinase C phosphorylation site"
FT	376. .382	FT	765. .767
FT	Region	FT	/note= "Protein kinase C phosphorylation site"
FT	/note= "MHC immunoglobulins and major histocompatibility complex proteins signature"	FT	807. .810
FT	398. .401	FT	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	FT	809. .811
FT	426. .428	FT	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	FT	812. .815
FT	453. .455	FT	/note= "cAMP and cGMP-dependent protein kinase phosphorylation site"
FT	Modified-site	FT	841. .844
FT	467. .470	FT	/note= "Amidation site"
FT	Modified-site	FT	865. .867
FT	467. .469	FT	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	FT	898. .903
FT	475. .477	FT	/note= "N-myristoylation site"
FT	Modified-site	FT	911. .913
FT	481. .484	FT	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	FT	952. .954
FT	486. .491	FT	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	FT	965. .967
FT	492. .495	FT	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	FT	980. .982
FT	493. .498	FT	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	FT	1031. .1038
FT	515. .517	FT	/note= "Tyrosine kinase phosphorylation site"
FT	Modified-site	FT	1034. .1036
FT	527. .530	FT	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	FT	1038. .1041
FT	532. .535	FT	/note= "Amidation site"
FT	Modified-site	FT	1067. .1070
FT	546. .548	FT	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	FT	1103. .1105
FT	552. .557	FT	/note= "Protein kinase C phosphorylation site"
FT	560. .563	FT	1120. .1122
FT	Modified-site	FT	
FT	561. .563	FT	

/note= "protein kinase C phosphorylation site"

FT XX WO200123585-A1.
XX XX 05-APR-2001.
XX PD 29-SEP-2000; 2000WO-US026915.
XX PF 30-SEP-1999; 99US-00408865.
XX PR (MILL-) MILLENNIUM PHARM INC.
XX PA Kapeller-Libermann R;
XX PI WPI; 2001-235374/24.
XX DR N-PSDB; AAD03459.
XX DR
XX XX Nucleic acids encoding human ubiquitin protease polypeptides, useful for
PT PT preventing, diagnosing and treating, e.g. breast, lung or liver cancer
PT PT and viral infections.
XX XX
XX PS Claim 9; Fig 1; 120pp; English.
XX CC The present sequence is human ubiquitin protease belonging to the family
CC of mammalian deubiquitinating enzymes. This ubiquitin protease is encoded
CC by the cDNA insert of plasmid deposited with ATCC NO: FTA-1849. Ubiquitin
CC protease is highly expressed in foetal kidney, testes, foetal liver, and
CC ovary and foetal heart. It is also expressed in liver, breast, lung, and
CC colon tissue and in liver metastasis derived from malignant colon
CC tissues. Ubiquitin protease is useful for treating disorders mediated by
XX XX
XX XX Query Match 35.0%; Score 1423.5; DB 4; Length 1123;
XX XX Best Local Similarity 37.3%; Pred. No. 2.8e-102;
XX XX Matches 347; Conservative 112; Mismatches 238; Indels 233; Gaps 25;
XX XX
XX QY 1 MTIVDKASESDSFAYQNGQSSSEAVSPGMDAGSAGVSSINDV----- 47
XX DB 1 MPIVDKLKEAL-----KPRKDSADGDELKILLASSAKVLLQKTEFFPASKSFYQL 53
XX QY 48 ----SNHTLSLGFVPVAVVY-SSSSVPDKSPQKQDQALGDGIAPPKQVLPFSEKICLK 102
XX DB 54 EALKSKVLLNPKTEGASRHSKGGDDPARQSGSEHTYESCGDVPAPQKVLFPPTLSLR 113
XX QY 103 WOQTHRVGAGLQNTGTCFANAALQCLTYTPPLANMLSEHSKTCBAEGFCMCTWQAH 162
XX DB 114 WERFVRVAGLHNLGNTCFNLNATIQCLTYTPPLANMLSEHSKTCBAEGFCMCTWQAH 173
XX QY 163 ITQALSNPGDVIKPMFVINEMRITARHLRFQNGQDAHEFLQYTDAMOKACLSNGKILDR 222
XX DB 174 IVQAFANSAGNAIKPVSFIRDLUKIARHFRFQNGQDAHEFLRYTTIDAMOKACLSNGKILDR 233
XX QY 223 HTQATTLVCQIFGGYLSRVKCLNCKGVSDTFDPLDTLLEIKAAQSVNKALBOFVKPEQ 282
XX DB 234 QTQATTLVHQIFGGYLSRVKCVCKSVSDTYDPLDVALEIRQAANIVRALELFPKADV 293
XX QY 283 LDGNSYKCKCKKQVVPASKRFTTHRSNVLTLKREANFTGGKIADKVKYVPELDRP 342
XX DB 294 LSGENAYMCAKCKKVPASKRFTTHRSNVLTLKREANFTGGKIADKVKYVPELDRP 353
XX QY 343 YMSQNGPEPIVYVAVLVHTGFCNCHAGYFCYIKASNGLWQMNDSIVSTSDIRSVLSQ 402
XX DB 354 YMSQNGPEPIVYVAVLVHTGFCNCHAGYFCYIKASNGLWQMNDSIVSTSDIRSVLSQ 413
XX QY 403 QAYVLYIRSHDVKNNGE----LTHPTSPQSSSRP-----VLSQVNTNQ-- 446
XX DB 414 QAYVLYIRSHDVKNNGE----LTHPTSPQSSSRP-----VLSQVNTNQ-- 473
XX QY 447 -----AAGFTGQPLPS-----HMTKNPHLNGTGPLKD 475
XX DB 474 GTMKPKHTTEIGVPISRNGSTGLKSGNCGICPPKLPSPGSPKUSQUTHEM-----PTILD 530
XX QY 476 TPSSSMSSP-----NGNSS-----VNRASPV----- 496

Db 531 DPGKKVKKPAPPQHFSPRTAQGLPCTSNSSRSRSGSQGWSRDVULSTSPKLLATAT 590
QY 497 -----NASASYQNVNVRSSVPIPEHPK-----KQKITISI--H 527
Db 591 ANGHGLKGNDESAGLDRGSSSSSSPEHSASDSTKAPQTPRSGAAHLCDSQETNCSTAGH 650
QY 528 NKLVPVRCQSQ-----PNLHNSNLE-----NPT 550
Db 651 SKTPPGGADSKTVLKSPVLSNNTTTEPASTMSPPPAKKLALSAKKASTLWRATGNDLRPP 710
QY 551 KPVPSTTNSAVQSTSN---ASTMSV-SSKYTKPIPRSESCSQPVNMGKSKLNSVLVP 606
Db 711 PPSPPSGLTHP--MKTSHFPVAVSTWVHRARVAPQSSSRQLQPPSPHPILLSTPKP 768
QY 607 YGAESSEDSDESKGLGKENGIGITVSSHSPGQDAEDEATPHELOEPMTLNAGANSADSD 666
Db 769 PGTSEPRSSCSISTA..PQVN-----EDLVSLPHQLPEA----- 801
QY 667 SDPKENGLAPDGASQXQGP-ALHSENPFKANGLPKMLPAPLLSLPEDKILETFRLSNK 725
Db 802 SEPPQSPSEKSKTKTVGEPQLGSE-----TLRPOHIREA--TAAPHGK-----RKRK 847
QY 726 LKSGTDEMSAPGAER3PPEDRDAEPQGPSP 755
Db 848 KKCRPEDTAASALQES-----QTQROFGSP 872
RESULT 10
AAB93200
ID AAB93200 standard; protein; 954 AA.
XX AC AAB93200;
XX XX 26-JUN-2001 (first entry)
XX DE Human protein sequence SEQ ID NO:12164.
XX KW Human; primer; detector; diagnosis; antisense therapy; gene therapy.
XX OS Homo sapiens.
XX XX EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-00116126.
XX PR 29-JUL-1999; 99JP-00248036.
XX PR 27-AUG-1999; 99JP-00300253.
XX PR 11-JAN-2000; 2000JP-00118776.
XX PR 02-MAY-2000; 2000JP-00183767.
XX PR 09-JUN-2000; 2000JP-00241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX DR WPI; 2001-318749/34.
XX XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX PS Claim 8; SEQ ID NO 12164; 2537pp + Sequence Listing; English.
XX CC The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence and the 3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 954 AA;

Query Match 35.0%; Score 1422.5; DB 4; Length 954;
Best Local Similarity 41.1%; Pred. No. 2.6e-102; Mismatches 244; Indels 127; Gaps 25;
Matches 335; Conservative 110; Mismatches 244; Indels 127; Gaps 25;

QY 1 MTIVDKASSESDPSAYQNGPGSEAVSPGMDAGSAGWAVSSINDV----- 47
Db 1 MPVIDKLKEAL-----KPGKDSADDGELGKLASSAKKVLQKIEPEPASKSPSYQL 53

QY 48 ----SNHTLSLGPVGA VVY-SSSVDPKSPSPQDQALGDGIAPQKVLPSEKICLK 102
Db 54 EALKSKYVLLNFKETEGASRHKSGDDPPARQSGEHTYESCGVPAPOKVLPTTERLSLR 113

QY 103 WQOTHRVGAGLQNLGNTCFANAALQCLTTPPLANTMLSHESKTCARSGFCMCTMQAH 162
Db 114 WERFVRVGAGLHNLGNTCFNLATIQCLTTPPLANTMLSHESKTCARSGFCMCTMQAH 173

QY 163 ITQALSNPQDVTKMPVINEMRRIARHFRFGQEDAEHEFLQYTVDMQKACLINGSNKLDK 222
Db 174 IVQAFANGSNAIKPVSFIRDLKRIARHFRFGQEDAEHEFLRYTIDAMQKACLINGCAKLDK 233

QY 223 HTQATTLVCQIFGGYLSRVKCLNCKGVSDTDPDYLDITLEIKAAQSVNVALEQVFKPEQ 282
Db 234 QTOATTLVHQIFGGYLSRVKCLNCKGVSDTDPDYLDITLEIKAAQSVNVALEQVFKADV 293

QY 283 LDGENSYKSCCKKMPVASKRFTIHRSSNVLTLSLKRFAFTGKGIADVKYPEYLDIRP 342
Db 294 LSGENAYVCAKCKKVPASKRFTIHRSSNVLTLSLKRFAFTGKGIADVKYPEYLDIRP 353

QY 343 YMSQNGEPIVYLVYVHTGFCNCHAGYCYTKASNGLYQMNDISVTSDIRSVLSQ 402
Db 354 YMSQNGDPVMYGLYVAVLVHSGYCHAGYCYTKASNGLYQMNDISVHSSNVKVLNQ 413

QY 403 QAYVLFYTRSHDVKNKGGLTHTPSG-QSSPRPVISORVTVNQAPGFTGPQ-LPSHM 460
Db 414 QAYVLFYLR-----IPGSKSPEGLISR---TGSSSLFG--RPSVIPDHS 453

QY 461 IKNPPLHNGTGL--KDTPTSSMSPPNGSNVNRASPVNASQVMSVNRSSVIBEPK 518
Db 454 KKNIGNGIISPLTGKQDSGTMKPHTEI--GVPIRNGSTLGLKQNGCIPPKLPS 511

QY 519 KQKTIISHNKLPRVQCQSNLHNSLENP-----TKPVPS---STITNSAVQSTNSAT 571
Db 512 GS-----PSFKLSQPTHTFTILDDPGKKVKKPAPQPHSPRTAQGLPCTSN 560

QY 572 MSVSSKVTKPIPRSESCSQPMNCKSLNSVLV-PVGAESSESDSDESKGLGKENGIGT 630
Db 561 ---SSRSGSORQSGWRDVLVSTSPKLLATATANGHLK-----GNDESAGLDR-RGSSS 612

QY 631 IVSSHSPQDAEEDAEATPHELQEPMTLNGANSADSDPKENGLAPDGACQCPALHSE 690
Db 613 SSPEHSASSDSTRAQPFERS-----GAHLCDQSE-----TNC--STAGHSK 652

QY 691 NPAKANGLPGKLM-----PAPLLSLPEDKILETFLSNKLGSTDEMSAPGAER 740

Db 653 TPPSGADSKTVKLKSPVLSNNTTTEPASTMGPPAK-----KLALSAKKASTLWRATG--- 704
QY 741 GPPEDRDAEPQPGSPAASLEEP-----DAAASLFP 771
Db 705 -----NDLRPPFPSPSSD-LTHPMKTSHPVVASTWP 734

RESULT 11
AAB95208
ID AAB95208 standard; protein; 548 AA.
XX
AC AAB95208;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:17313.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
PT
PT
XX
PS Claim 8; SEQ ID NO 17313; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 548 AA;

Query Match 33.8%; Score 1375; DB 4; Length 548;

	Best Local Similarity	47.3%;	Pred No. 6.1e-99;	Mismatches	81;	Indels	86;	Gaps	12;
1	MTIVDKASSSSDPFSAYQNOGSPGSEAVSPGDDMDAGSAGCAVSSINDV	---	---	---	---	---	---	---	47
1	MPITVDVXLKEAL	-----	KPGRKDSADGELGKGLASSAKVLLQKIEFEFPASKSFSYQL	53					
48	---	---	---	---	---	---	---	---	---
54	EALKSKVILLNPKTEGASRHKSGDDPPARRRGSEHTYESGDBGVPAQKVLFPFTERLSLR	113							
103	WQOTHRVRGAGLQNLGNTCFANALQCLITTPPLANYMLSHEHSKTCBAEGECMMCTMOAH	162							
114	WERVFRVRGAGLHNLGNTCFLNATIQCLITTPPLANYLLSKEHARSCHOGSPCMLCVMQNH	173							
163	ITCALSNPGDVIKPMFVFINEMRRIARHLRFNGQDAHEFLQYTVDMQKACLINGSNKLDLR	222							
174	IVQAFANGNAIKPVSFIRDLKKIARHFRFGQDAHEFLRYTIDAMQKACLNGCAKLDLR	233							
223	HTQATTLVCOIFGGYILRSRVKCLNCKGVSDTDPDYLDITLTKAOSVNKALEQVFVKPEQ	282							
234	QTQATTLVHQIFGGYILRSRVKCSVKSVDYDPDYLDVALETIRQANIVRALELFPKADV	293							
283	LDGENSKYCSCKKQWVPASKRFTIHRSSNVLTLSLKRFRANFTGGKIADKVKYPEYLDIRP	342							
294	LSGENAYMCAKCKKKVPASKRFTIHRSSNVLTLSLKRFRANFSGGKITKDVGYPEFLNIRP	353							
343	YMSQPNCEPIVYLYAVLHVTGNCCHAGHYCYCIKASNLGLWQMNDSIVSTDGIRSVLSQ	402							
354	YMSQNGDPMYGLYAVLHVSIGSCHIAGHYCYIVKASNGQWQMNDSLHSSNVKVLNQ	413							
403	QAYVLFIIRSHDVKNNGGELTHPTHSFG--QSSPRPVIISQRVVTNKQAPFGIQ-LPSIMH	460							
414	QAYVLFIYLR-----	-----	IPGSKSKPEGLISR--	TGSSSLPG--	RPSVLPDHS	453			
461	IKNPPLHNGTQPL--	KDTPSSMSXPNGNSVNRASFPVNASVQWNSVNRSSVIEPEHK	518						
454	KKNIINGIIISPLTGRQDGTWKKPHTTBEI--	GVPIISRNGTILGLKSONGCIIPKALPS	511						
519	KQKITISIHNLKPLVRQCSQPNLHNSLENPTKVPSPSITITNSAVQSTSNASTMSVSSKV	578							
512	GS-----	---PSPKLQSTPHTMPTILDDPGK-----	---	---	---	---	---	---	---
579	TKPIIP	583							
537	KKPAP	541							

RESULT 12	
ADC27023	
ID	ADC27023 standard; protein; 565 AA.
XX	
AC	ADC27023;
XX	
XX	18-DEC-2003 (first entry)
DT	
XX	
DE	Human deubiquitinating enzyme DUB4.8 SEQ ID NO:20.
XX	
XX	human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
KW	infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
XX	
XX	
OS	Homo sapiens.
XX	
PN	WC02003072724-A2.
XX	
PD	04-SEP-2003.
XX	
XX	20-FEB-2003; 2003WO-US005338.
XX	
PR	22-FEB-2002; 2002US-0358873P.
PR	22-FEB-2002; 2002US-0358875P.
PR	08-MAR-2002; 2002US-0363020P.
PR	12-APR-2002; 2002GB-00008404.
PR	

XX	(AVET) AVENTIS PHARM INC.
PA	
XX	Hahn C, Liu H;
PI	
XX	WPI; 2003-721759/68.
DR	N-PSDB; ADC27022.
DR	
XX	New isolated polynucleotide encoding human deubiquitinating proteases,
PT	useful for identifying e.g. antiinflammatory or immunosuppressive agents,
PT	also the encoded polypeptides.
PT	
XX	Claim 2; SEQ ID NO 20; 201pp; English.
PS	
XX	The present invention describes human deubiquitinating protease enzymes
CC	(I), designated hDUB.
CC	The enzymes designated hDUB4 are located on
CC	chromosome 4, and the enzymes designated hDUB8 are located on chromosome
CC	8. Also described is a method for reducing inflammation, modulating
CC	autoimmune diseases and modulating immune reactions during infections by
CC	administering a compound (A) that inhibits (I) or alters regulation of
CC	the transcription of nucleic acid sequences encoding (I). (I) has
CC	antiinflammatory and immunosuppressive activities, and can be used in the
CC	regulation of cytokine/cytokine receptor signaling involved in lymphocyte
CC	proliferation. (I) can be used for identifying their respective
CC	inhibitors (A). (A) are useful for reducing inflammation and to modulate
CC	autoimmune diseases or immune reactions during infections. The present
CC	sequence represents hDUB4.8 from the present invention.
XX	
SO	Sequence 565 AA;
XX	

Query Match	28.3%;	Score 1150.5;	DB 7;	Length 565;
Best Local Similarity	46.7%;	Pred. No. 2.8e-81;		
Matches 244;	Conservative	78;	Mismatches 162;	Indels 39;
				Gaps 8

Qy	28	PGDMDAGS	AWGAVS	SLNDVSNH	TLTSLG	CPVCAV	VVYSSSS	VPDKSP	QPKDQAL	GDGIA	87																						
Db	33	PVDMED	SLYL	UGGWFQ	NHFSKLT	-SSRP	DAAFAEI	QRTSL	PEKSP	LCSETR	VDLCCDLA	91																					
Qy	88	PPQKVL	FPSEK	ICLFL	FWOOTH	RVRG	AGLQNL	GNCTC	PANAA	LQCL	TYTPP	PLAN	YML	SHHSKT	147																		
Db	92	PVARQL	APRE	KLPL	-SRR	PAAV	AGAGL	QNMGN	TCVYN	ASLO	CLTYTPP	PLAN	YML	SHHSQ	151																		
Qy	148	CHAEFG	CMQCT	MOAHI	TOALSN	PGDV	IKPMF	VINEM	RRIR	AHLR	FGNO	DEAHE	FLQY	TVD	207																		
Db	152	CHRHK	GCMLC	TMOAHI	TRALHN	PGHVI	QP-----	SQALA	AGFHR	GKQ	DEAHE	FLM	TVD	205																			
Qy	208	AMQAC	ILNGS	NKLDR	HTQAT	TLV	COIF	GGYLR	SRVK	CLNC	KGV	SDTF	PYLD	ITL	EIKAA	267																	
Db	206	AMKKA	CLPGH	KQVD	HDHS	KD	TTLIHQ	IFGGY	WR	SQIK	LCH	CHGIS	DST	DFPYLD	IALD	IOAA	265																
Qy	268	QSVN	KALQ	FOF	VKE	PEOL	DGEN	SYK	SKCK	KWVP	PASK	RFT	THRS	SNV	LTLSL	KRFAN	FTGGK	327															
Db	266	QSVQ	ALBQL	VKE	PEL	INGEN	AYHC	GVCL	QRA	PASK	TLT	HTSA	KVL	LILV	LK	KRFS	SDVTG	GNK	325														
Qy	328	IAKDV	KYPE	YLD	IR	YMS	QNG	PE	VYV	LYAV	LVT	FGN	CHAG	HYFC	YIKAS	NGLW	QYWN	387															
Db	326	IAKN	QY	PE	CLDM	QYMS	QO	NT	G	PLV	YV	LYAV	L	VHAG	WSCH	NGHY	FYS	VKA	EGOW	YKMD	385												
Qy	388	DSIV	TS	DIR	SV	LVLS	QOAY	VLF	VIIR-----	SHD	VKN	GGEL	TH	PTHPS	PCQS	SPRP	VI	SQR	440														
Db	386	DAE	VTAA	SITS	VLVS	QOAY	VLF	YI	QKSE	WHER	SES	VR	GRHE-----	PRAL	GAED	433																	
Qy	441	VVTN	QOAA	PGFT	------	GPQL	F	SHMI	KNP	PHLM	GT	GPL	KD	TPSS	SSM	SPNGN	-SSVN	491															
Db	434	--TD	RAT	OGEL	KR	DP	CL	QOAF	DEL	HL	VER	AT	OEST	L	DH	WK	FLO	BQNK	TK	FE	NKVR	KVE	491										
Qy	492	RAS	P	VNAS	AS	VQ	NWS	VNR	SVI	PE	HP	KQ	KI	TI	S	H	N	K	L	P	V	T	Q	534									
Db	492	GTLP	--PD	V	L	V	T	H	Q	S	K	Y	K	C	K	M	K	N	H	P	E	Q	S	S	L	N	-SS	T	P	T	H	Q	532

RESULT 13
ABU10207

CC proliferation. (I) can be used for identifying their respective
 CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
 CC autoimmune diseases or immune reactions during infections. The present
 CC sequence represents hDUB4.6 from the present invention.

XX
 SQ Sequence 530 AA;

Query Match 28.1%; Score 1142.5; DB 7; Length 530;
 Best Local Similarity 46.7%; Pred. No. 1.1e-80;
 Matches 243; Conservative 77; Mismatches 161; Indels 39; Gaps 8;

QY 271 NKALEQFVKPEQLDGENSYKSKCKMWPASKRFTTHRSSNVLTSLKRFANFTGGKIAK 330
 DB 234 QQALEQLVKPEELNGENAYHCGVCLQAPASKTLTHTSAKVLILVLRFSVDVTGNKIAK 293

QY 331 DVKYPEYLDIRPYMSQNGEPIVYVLYAVLVHTGCHAGHYFCYIKASNGLWYQMNDISI 390
 DB 294 NVQYPECLDMQPYMSQQTGTPLVYVLYAVLVHAGWSCHNGHYFSYVKAQEGWYKMDAE 353

QY 391 VSTDIRSVLSQOAYVLYFYIR-----SHDVKNKGELTHPTSPGQSSPRPVISORVVT 443
 DB 354 VTASSITSVLSQOAYVLYFYIQSEWERHSESVSRRGE-----PRALGAED--T 399

QY 444 NKQAAFGFI-----GPOLPSHMIKNPPHNLGTGPKLDTSPSSMSSPNGNSVNRSAP 495
 DB 400 DRRATQGLKRDHPCLQAPDELHVE-----RATQESTLDHMKFLQEQNKTKP 448

QY 496 VNASASVQWNSVR-----SSVPIPHKPKQKITISIHKNLPRCOCOPNLHNSLENP 549
 DB 449 -----EFNVRKVEGTLPDPVLVTHQSKYKCGMKNHHP-----EQSSLLNLSSTNP 494

QY 550 TKPVPSSTITNSAVOSTSNASTM-SVSSKVKTKPIPRSESCSQPVM 593
 DB 495 TD-----QESMNTGILASLQGRTRRAKGNKHKCRSLL 527

RESULT 14
 ADC27019
 ID ADC27019 standard; protein; 530 AA.
 XX
 AC ADC27019;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human deubiquitinating enzyme DUB4.6 SEQ ID NO:16.
 XX
 KW human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
 XX infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
 XX
 OS Homo sapiens.
 XX
 PN WC2003072724-A2.
 XX
 PD 04-SEP-2003.
 XX
 PF 20-FEB-2003; 2003WO-US005338.
 XX
 PR 22-FEB-2002; 2002US-0358873P.
 XX
 PR 22-FEB-2002; 2002US-0358875P.
 XX
 PR 08-MAR-2002; 2002US-0363020P.
 XX
 PR 12-APR-2002; 2002GB-00008404.
 XX
 XX (AVET) AVENTIS PHARM INC.
 XX
 XX Hahn C, Liu H;
 XX
 XX WPI; 2003-721759/68.
 XX
 DR N-PSDB; ADC27018.
 XX
 PT New isolated polynucleotide encoding human deubiquitinating proteases,
 XX useful for identifying e.g. antiinflammatory or immunosuppressive agents,
 XX also the encoded polypeptides.
 XX
 XX Claim 2; SEQ ID NO 16; 201pp; English.
 XX
 XX The present invention describes human deubiquitinating protease enzymes
 CC (I), designated hDUB. The enzymes designated hDUB4 are located on
 CC chromosome 4, and the enzymes designated hDUB8 are located on chromosome
 CC 8. Also described is a method for reducing inflammation, modulating
 CC autoimmune diseases and modulating immune reactions during infections by
 CC administering a compound (A) that inhibits (I) or alters regulation of
 CC the transcription of nucleic acid sequences encoding (I). (I) has
 CC antiinflammatory and immunosuppressive activities, and can be used in the
 CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte

CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
 CC autoimmune diseases or immune reactions during infections. The present
 CC sequence represents hDUB4.6 from the present invention.

XX
 SQ Sequence 530 AA;

Query Match 28.1%; Score 1142.5; DB 7; Length 530;
 Best Local Similarity 46.7%; Pred. No. 1.1e-80;
 Matches 243; Conservative 77; Mismatches 161; Indels 39; Gaps 8;

QY 31 MDGASGAGVSSLMNVNHTLSLGPVCAVYSSVPSDKSPQKQDQALGDGIAPQ 90
 DB 1 MEDDSLVLGGEGQNFHFKLT--SSRPDAFAFIQRTSLPEKSPSCETRVDCDDLPVA 59

QY 91 KYLFSEKICLKWQCTHRVAGLQNLGNTCFANAALQCLTYTPPLANYMLSHHSKTHA 150
 DB 60 RQLAPREKLPSSRRPAVAGLQNNGNTCYNASLQCLTYTPPLANYMLSHHSKTHA 119

QY 151 EGFQNMCTMOAHITQALSNPQGVKIPMEVINEMRRIARHLRFGNOEDAEHFLQYTVDMQ 210
 DB 120 HKGCMCTMQAHTTRALHNPQHVQP-----SQALAGFHRGKQDAHEFLMTVDAMK 173

QY 211 KACLNGSNKLDHRTQATTLVCOIFGYSLSRVKCLNCKGVSTFDPYLDITLEIKAAQSV 270
 DB 174 KACLPGHKQVDHHSKJOTLIHQIFGYSWSQIKCLHCHGISDTFDPYLDIADLQAAQSV 233

QY 271 NKALEQFVKPEQLDGENSYKSKCKMWPASKRFTTHRSSNVLTSLKRFANFTGGKIAK 330
 DB 234 QQALEQLVKPEELNGENAYHCGVCLQAPASKTLTHTSAKVLILVLRFSVDVTGNKIAK 293

QY 331 DVKYPEYLDIRPYMSQNGEPIVYVLYAVLVHTGCHAGHYFCYIKASNGLWYQMNDISI 390
 DB 294 NVQYPECLDMQPYMSQQTGTPLVYVLYAVLVHAGWSCHNGHYFSYVKAQEGWYKMDAE 353

QY 391 VSTDIRSVLSQOAYVLYFYIR-----SHDVKNKGELTHPTSPGQSSPRPVISORVVT 443
 DB 354 VTASSITSVLSQOAYVLYFYIQSEWERHSESVSRRGE-----PRALGAED--T 399

QY 444 NKQAAFGFI-----GPOLPSHMIKNPPHNLGTGPKLDTSPSSMSSPNGNSVNRSAP 495
 DB 400 DRRATQGLKRDHPCLQAPDELHVE-----RATQESTLDHMKFLQEQNKTKP 448

QY 496 VNASASVQWNSVR-----SSVPIPHKPKQKITISIHKNLPRCOCOPNLHNSLENP 549
 DB 449 -----EFNVRKVEGTLPDPVLVTHQSKYKCGMKNHHP-----EQSSLLNLSSTNP 494

RESULT 15
 ADC27015
 ID ADC27015 standard; protein; 530 AA.
 XX
 AC ADC27015;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human deubiquitinating enzyme DUB4.3 SEQ ID NO:12.
 XX
 KW human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
 XX infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
 XX
 OS Homo sapiens.
 XX
 PN WO2003072724-A2.
 XX
 PD 04-SEP-2003.
 XX
 PF 20-FEB-2003; 2003WO-US05338.
 XX
 PR 22-FEB-2002; 2002US-0358873P.
 XX
 PR 22-FEB-2002; 2002US-0358875P.
 XX
 PR 08-MAR-2002; 2002US-0363020P.
 XX
 PR 12-APR-2002; 2002GB-00008404.

```
XX (AVET ) AVENTIS PHARM INC.
XX Hahn C, Liu H;
XX WPI; 2003-721759/68.
XX N-PSDB; ADC27014.
XX New isolated polynucleotide encoding human deubiquitinating proteases,
XX useful for identifying e.g. antiinflammatory or immunosuppressive agents,
XX also the encoded polypeptides.
XX Claim 2; SEQ ID NO 12; 201pp; English.
XX
XX The present invention describes human deubiquitinating protease enzymes
XX (I), designated hDUB. The enzymes designated hDUB4 are located on
XX chromosome 4, and the enzymes designated hDUB8 are located on chromosome
XX 8. Also described is a method for reducing inflammation, modulating
XX autoimmune diseases and modulating immune reactions during infections by
XX administering a compound (A) that inhibits (I) or alters regulation of
XX the transcription of nucleic acid sequences encoding (I). (I) has
XX antiinflammatory and immunosuppressive activities, and can be used in the
XX regulation of cytokine/cytokine receptor signaling involved in lymphocyte
XX proliferation. (I) can be used for identifying their respective
XX inhibitors (A). (A) are useful for reducing inflammation and to modulate
XX autoimmune diseases or immune reactions during infections. The present
XX sequence represents hDUB4.3 from the present invention.
XX
XX Sequence 530 AA;
XX
Query Match 28.0%; Score 1140.5; DB 7; Length 530;
Best Local Similarity 46.5%; Pred. No. 1.5e-80;
Matches 242; Conservative 78; Mismatches 161; Indels 39; Gaps 8;
Qy 31 MDAGSASGAVSSLDVSNHTLSLGPVPGVAVVYSSSVDPKSPQKQDQALGDGIAPPQ 90
Db 1 MEDDSLVLGGEWQHFESKLT-SRPDAFAEIQTSLPEKSPSLCETRVLDLDDLPVA 59
Qy 91 KVLFPSEKICLKWQOTHRVAGLQNLGNCTCFANAALQCLTYTPPLANMLSHEHSKTC 150
Db 60 ROLAPREKLPLSSRRPAAGVAGLQNMGNCTCVNASLQCLTYTPPLANMLGREHSQTC 119
Qy 151 EGFQCMCTMQAHITQALNPGDVLKMPFVINEMRRIARHFRGQEDAEHFLQTVDDAQ 210
Db 120 HGGCMCTMQAHITRALNPGHVLP-----SQALAGFRGQEDAEHFLMFTVDAMK 173
Qy 211 KACLNGSNKLDRTQATTLVCQIFGYYLRSRVKCLNCKGVSDFDPYLDITLTKAAQSV 270
Db 174 KACLPGHKQVDHSHKDTLHIFGYYWRSQIKCLHCHGISDTPPYLDIALDICAQSV 233
Qy 271 NKALBQFVKPQLDGENSYKCSKCKMVPASKRFTIHRSSNVLTLSLRPFANFTGGKIAK 330
Db 234 QQALEQLVKPEELNGENAYHCGVCLQAPASKTLTHTSAKVLILVLRFSDDVTGNKIAK 293
Qy 331 DVKYPEYLDIRPYMSQNGEPIVVYLVAVLVHTGPNCHAGHYECYIKASNGLWYQMDSI 390
Db 294 NVQYPECLDMQPYMSQNGTGPLVYLVAVLVHAGWSCHNGHYFSYVRAQEGQWYKMDAE 353
Qy 391 VSTDIRSVLSQQAYLVFYIR-----SHDVXNGGELTHTSPGQSSRPVTSQRVVT 443
Db 354 VTAASITSVLSQQAYLVFYIQSEWERHSESVSRGRE-----PRALGAED--T 399
Qy 444 NKQAPGFI-----GPQLPSHMKNPPLHNGTGPLKDTFPSSSMSSPNGN-SSVNRAS 494
Db 400 DRRATQGLKRDHPCLQAPLDELHVERATQESTLDRWKFLQEQNKTKPEFNVKRGVGL 459
Qy 495 PVNASASVQNWSVNRSSVYPEHPKKQKITITSHNKLVPVQ 534
Db 460 P--PDVLVIHQSKYKCGMKNHHPQQSSLLNLSSTPTHQ 497
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:56:53 ; Search time 20 Seconds
(without alignments)
3727.422 Million cell updates/sec

Title: US-10-049-745-4
Perfect score: 4070
Sequence: 1 MTIVDKASESDPSAYQNQP.....AAESLEPDAAASLFPFSEG 775

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues 283366
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	846.5	20.8	526	2	JC6133	deubiquitinating e
2	760	18.7	457	2	T41069	ubiquitin carboxyl
3	720	17.7	891	2	H86306	F20D23.20 protein
4	703	17.3	743	2	B84639	probable ubiquitin
5	692.5	17.0	631	2	T05103	hypothetical prote
6	626.5	15.4	1008	2	T05578	hypothetical prote
7	504.5	12.4	1082	2	T05075	hypothetical prote
8	428	10.5	471	2	S57591	ubiquitin specific
9	370.5	9.1	557	2	T50788	ubiquitin carboxyl
10	358.5	8.8	449	2	T37611	hypothetical prote
11	351.5	8.6	1672	2	T46237	hypothetical prote
12	343.5	8.4	792	2	S63141	probable ubiquitin
13	336	8.3	365	2	B84611	ubiquitin-specific
14	332	8.2	805	2	S50277	deubiquinating enz
15	323.5	7.9	371	2	T06097	hypothetical prote
16	304	7.5	926	2	S39344	hypothetical prote
17	303.5	7.5	264	2	T47183	probable ubiquitin
18	296	7.3	408	2	S40715	hypothetical prote
19	295.5	7.3	585	2	T50325	hypothetical prote
20	287.5	7.1	825	2	T47164	hypothetical prote
21	275	6.8	555	2	T23531	ubiquitin carboxyl
22	267	6.6	438	2	T39772	protein T05H10.1
23	265.5	6.5	1292	2	A98225	hypothetical prote
24	265.5	6.5	1292	2	T24559	fat facets protein
25	264.5	6.5	2559	2	T30850	fat facets (faf) s
26	260	6.4	2747	2	B49132	probable ubiquitin
27	257	6.3	512	2	T40815	hypothetical prote
28	256	6.3	754	2	S50601	hypothetical prote
29	254	6.2	1430	2	T21910	hypothetical prote

30	253	6.2	849	2	T41085	probable ubiquitin
31	251	6.2	1108	2	T41188	probable ubiquitin
32	250.5	6.2	1230	2	S53974	hypothetical prote
33	246.5	6.1	928	2	T04192	hypothetical prote
34	243.5	6.0	937	2	T04194	probable ubiquitin
35	243	6.0	350	2	T39795	hypothetical prote
36	242.5	6.0	1095	2	T20528	hypothetical prote
37	241	5.9	1130	2	T23104	hypothetical prote
38	241	5.9	1133	2	T23103	CDS protein F9L11.
39	236.5	5.8	887	2	C86453	hypothetical prote
40	229.5	5.6	376	2	T30979	hypothetical prote
41	229	5.6	794	2	I58376	hypothetical prote
42	219	5.4	688	2	S45803	ubiquitin thiolest
43	218	5.4	963	2	T09478	ubiquitin-specific
44	217	5.3	912	2	B44450	probable ubiquitin
45	217	5.3	914	2	T00757	

ALIGNMENTS

RESULT 1

JC6133
deubiquitinating enzyme - mouse
N:Alternate names: DUB-1 protein
C:Species: Mus musculus (house mouse)
C>Date: 16-Apr-1997 #sequence_revision 09-May-1997 #text_change 05-Nov-1999
C:Accession: JC6133
R:Zhu, Y.; Carroll, M.; Papa, F.R.; Hochstrasser, M.; D'Andrea, A.D.
Proc. Natl. Acad. Sci. U.S.A. 93, 3275-3279, 1996
A:Title: DUB-1, a deubiquitinating enzyme with growth-suppressing activity.
A:Reference number: JC6133; MUID:96194957; PMID:8622927
A:Accession: JC6133
A:Molecule type: mRNA
A:Residues: 1-526 <ZHU>
A:Cross-references: GB:U41636; NID:g1302629; PID:AA052532.1; PID:g1302630
C:Comment: This enzyme is the first enzyme of the ubiquitin system directly implicated in and cell cycle progression, and in cytokine-induced cell proliferation.

Query Match 20.8%; Score 846.5; DB 2; Length 526;
Best Local Similarity 37.1%; Pred. No. 5.9e-43;
Matches 186; Conservative 89; Mismatches 152; Indels 75; Gaps 12;

QY	76	PKDQALGDIAPP-----QKVLFPSEKICLKWOQTHRVGAGLQNLGNTCFANAALQ	127
DB	8	PEADPPALSPDAPELHQDEAQQVVEELTVNGKHSLSWESQPGCGGLQNTGNSCYLNAALQ	67
QY	128	CLTYTPPLANYMLSHSHSCTHAEGFCMMCTMOAHITQAL--SNPGDVIKPMFVINEMRR	185
DB	68	CLTHTFPLADYMLSQSHSQTCCSPGCKLCAMEALVTQSLHSHSGDVMKPSHILTS---	124
QY	186	IARHLRFGNQEDAHEFLQYTVDAWQKACLNGSKLDRHTQATTLVQIFGGYLSRVKCL	245
DB	125	-AFKHK--QQEDAHEFLMTLETMHESCLQVHRQSKPTSEDSPIHDFIGGWRWSQIKCL	181
QY	246	NCKGVSDTFPPYLDITLLEIKAAQSVNKLAEQVFKPEQLDGENSVKSCCKKMPVPSKRFT	305
DB	182	LCQGTSDTYDFRIDIPLDITSSAQSVKQALWDTEKSELCDGNAYYCGCKQKWPASKTLH	241
QY	306	IHRSSNVLTSLKRFANFTGGKIAKQVYPEYLDIRPYMSQPNGEPIYVYVAVLVHTGF	365
DB	242	VHIAPKVLMVILNRFSAFTGNKLDKRVSYFEFLDKPFLSEPTGGPLPYALYAVLVHDGA	301
QY	366	NCHAGHYFCYIKASNGLWYQNDISIVTSDIRSVLSQAAYLVFIR-----SHDVKNG	418
DB	302	TSHSGHYFCVCKAGHKWYKMDDTKVTRCDVTSVLNENAYVLYVVOQANLKQVSDIMPEG	361
QY	419	--GELTHPTH-----SPQSSPRPVISO-----RWVTNKQ	446
DB	362	RINEVLDPYQLKKSRRKHKHKKSPFTEDLGEPCENRDKRAIKETSLGKGVLEVNHHK	421

A,Reference number: A84420; MUID:20083487; PMID:10617197
A,Accession: B84639
A,Status: preliminary
A,Molecule type: DNA
A,Residues: 1-743 <STO>
A,Cross-references: GB:AEO02093; NID:94572681; PIDN:AAD2
C,Genetics:
A,Gene: At2g24640
A,Map position: 2

Query Match	17.3%;	Score	703;	DB	2;	Length	743;
Best Local Similarity	31.2%;	Pred. No.	3.3e-34;				
Matches	189;	Conservative	92;	Mismatches	241;	Indels	84;
Gaps	15;						
QY	60	AVVSSSSVPDKSKSPSQKQALGDGIAPQKVLPFPSEKIC--LKWQOTHRVRVAGLQNLG	117				
DB	174	ASLPGNESASKTRIALVPQSQSKATLKPTDVLFPYESFVRYVYVNDRPTMAPCGLTNCG	233				
QY	118	NTCFANAALQCLTTPPLANYMLSHEHSKTCBAEGFCMKTQMAHITQALNSPGDVLKPM	177				
DB	234	NSCFANVVLQCLSWTRFLVAYLLERGHKRECRNDWCLCFEFENHLDRA--NYSRPFPSPM	292				
QY	178	FVINEMRRIARHLRFRGQEDAEHFLQVTVDAQKACLN--GSNKLDRHTQATTLVCOIF	234				
DB	293	NIISRLPNIGNLGYGQEDAEHMLRAIDMMQSVCLDFEGGEKVVPVRAQETTLIOYIP	352				
QY	235	GGYLRSRVKCLNCKGVSDTDPYLDITLLEIKA--AQSVNKALBQFVKPEQLDGENSYKSK	293				
DB	353	GGLLQSQVQCTACNSVSDQVENMMDLTVEIHGDAVSLSECLDQPTAKEWLQGDNLVKCDR	412				
QY	294	CKKMWPAKSKFTIHRSSNVLTSLSKRANFTGGKIAKDVKPYEYLDLRPYMSQPNGEPIV	353				
DB	413	CDQYVKKACKRLSRICANILTLAKRIFQGGFRGKLNKRISFETFDLGPYMSGGGESDV	472				
QY	354	VYLYAVLVHTGF--NCHAGHYFCYIKASNGLWYQNDPSIVSTDIRSLSQOAYVLYFIR	411				
DB	473	YKLYAVIVHLDMLNASFFGHYICVVKDFRGNWYRIDDSVEKVELEDVLQRAYMLLYSR	532				
QY	412	SHDVKNQGEIT-HPTHSPGGSSPPVVISORVVTNKQAAPGIGQLPSHMTKNPPLHNGT	470				
DB	533	T--VLSEGALTDDIWMYCRVQRP--	572				
QY	471	GPLKDTPTSSSMSSNGNSVNRASPVNANASVQNWNVNRSSVIEHPHKQKITTISHNKL	530				
DB	573	NTESNQDSVESSGVGTNDTSVSLNGLIIS-----HSEDPYEKESLSASV-----	620				
QY	531	PVRQ-----COSQPNL-----HSNLENPTKVPVPSSTTTNSAVQ	564				
DB	621	PVSEGEKEVDVKVTDVDSNRSRSDIMEHDSGTDHQEEFANGKEDP-----TVENLAVD	673				
QY	565	STSNASTMSVSSKVTKPIP-----RSECSQPVNMGKSKLNSSVLVPYGAESSEDDESK	620				
DB	674	SSCLDITTPSPSAETFIPOENERSDTSKPLEKHSDTSENKPL-----EKEHLDSESK	728				
QY	621	GLGKEN	626				
DB	729	PLEXEN	734				

RESULT 5

T05103
hypothetical protein F28M20.140 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05103
R:Bevan, M.; Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Hoheisel, J.; Mewes, F.
submitted to the Protein Sequence Database, November 1998
A:Reference number: Z15398
A:Accession: T05103
A:Molecule type: DNA
A:Residues: 1-631 <BEV>
A:Cross-references: EMBL:AL031004
A:Experimental source: cultivar Columbia; BAC clone F28M20

[illegible]

RESULT 6

T05578
hypothetical protein F22K18.240 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05578
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.
submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15419
A:Accession: T05578
A:Molecule type: DNA
A:Residues: 1-1008 <BEV>
A:Cross-references: EMBL:AL035356
A:Experimental source: cultivar Columbia; BAC clone F22K18
C:Genetics:
A:Map position: 4
A:Introns: 94/2; 521/3; 550/2; 579/1; 633/2; 674/3; 727/2; 755/3; 825/3; 84/
A:Note: F22K18.240

Db 257 LPNAKEYSRANNKQCECIHVHTFEGSLSSIVCPGQNNKTTIDPFDLSDLDIKKKKL 316
QY 271 NKALEQVKEPEQDGENSYKSKCKMVPASKRFTIHRSSNVLTSLKRFANFTGG---K 327
Db 317 YECLDSFHKKEQLKDFN-YHGECCNSTQDAIKQLGIHKLPSVLVQLKRFHEHLANGSNRK 375
QY 328 IAKDVVKPYEYLDIRPYMS-----QNGE--PIVYVLYAVLVHTGNCNCHAGHYFCVIKA 378
Db 376 LDDFIEPTTLNMKNYCTKEKDKHSENGKVPDIYIELIGIVSHKG-TVNEGHYIAFCKI 434
QY 379 SNGLWYQNDISVSTDIRSVLSQOAVLFY 409
Db 435 SGGWFRKFNDSMVSSISQEEVLKEQAVLLFY 465

RESULT 9

T50788
ubiquitin specific proteinase homolog - Arabidopsis thaliana
N;Alternate names: protein T30N20_60
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 02-Sep-2000
C;Accession: T50788
R;Byan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mew
submitted to the Protein Sequence Database, July 2000
A;Reference number: Z25240
A;Accession: T50788
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-557 <BEV>
A;Cross-references: EMBL:AL365234
A;Experimental source: cultivar Columbia; BAC clone T30N20
C;Genetics:
A;Map position: 5
A;Introns: 252/2; 292/1
A;Note: T30N20_60

Query Match 9.1%; Score 370.5; DB 2; Length 557;
Best Local Similarity 26.7%; Pred. No. 1.4e-14;
Matches 106; Conservative 66; Mismatches 148; Indels 77; Gaps 14;
QY 102 KWQOTHRVG-AGLNLGNTCFANAALQCLTTPPLANTYMLSHESK-TCHAEG---FCMM 156
Db 169 KW---TFPLGLRGLNLTGTCFNNVQLQALVHAPPLRNFWLSGQHNRDLCPRTWGLLCLP 226
QY 157 CTMOAHITQALSNPQDVVKPMFVINEMRRIARHLRFGNQEDAHEFLQYTVDM-----OK 211
Db 227 CDLDVFSAMFSGDRTPYSPAHLLYSWQHSTNLATYEQDSHEFFISLLDRIHENECKS 286
QY 212 ACLGSKNLDLRHTQATTVLCQIFGGYLSRVKCLNKGVSDFDPIYDITLIEI-----264
Db 287 KCLYQDNB-----ECQCITHKAFSGLLRSVDVCTTCGSTSTTYDPFIDISLTLDSMNGFS 341
QY 265 -----KAAQSVNK-----ALEOFVKPEQLDGENSYKSKCKMVPASKRFTIHR 308
Db 342 PADCKKNRYSGGPSVNAIMPTLSGCLDFTRSEKLGPOKLNQCSCGKRESSKQMSIRR 401
QY 309 SSNVLTLSLKEFANP-----TGKIAKDVVKPYEYLDIRPYM-----344
Db 402 LPLLLCLLVKRFESHLTRKTSRKIDSLQYPPRLANMSPYLSSSIIGKRFNGRIFAFDGE 461
QY 345 -----SQNGEPIVYVLYAVLVHTGNCNCHAGHYFCVIKASNGLWYQNDISVSTDIR 397
Db 462 EYDSSSSSSPSAE---FEIFAVVTHKGM-LESQVTVTLRL-KGLWYRCDDAWINEVEEE 516
QY 398 SVLSQOAVLVFIRSHDV-KNGGELTTPHSPQSSP 433
Db 517 VVRGCEYMLFYAQETVIOKAHKELSYQVISMADAPP 553

RESULT 10

T37611
ubiquitin carboxyl-terminal hydrolase - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T37611
R;Hunt, S.; Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, September 1995
A;Reference number: Z21730
A;Accession: T37611
A;Status: preliminary; translated from GH/EMBL/DBD
A;Molecule type: DNA
A;Residues: 1-449 <HUN>
A;Cross-references: EMBL:Z54096; PIDN:CAA90805.1; GSPDB:GN00066; SPDB:SPAC13A11.04c
A;Experimental source: strain 972h-; cosmid c13A11
C;Genetics:
A;Gene: SPDB:SPAC13A11.04c
A;Map position: 1
A;Introns: 32/3; 77/1; 306/2; 405/3; 427/2

Query Match 8.8%; Score 358.5; DB 2; Length 449;
Best Local Similarity 29.6%; Pred. No. 5.4e-14;
Matches 106; Conservative 62; Mismatches 143; Indels 47; Gaps 16;
QY 89 PQKVLFPSEKICLK-WQOTHRUGA-----GLNLGNTCFANAALQCLTTPPLANYMLSH 142
Db 120 PEKY---NQMVCLAEYKRPVPCATAGLRGTQNLGATCFMSVILQSLIHLNPLVRNLFESG 176
QY 143 EHSKTCHAEGFCMMCTMQAHITQALS--NPGDVIKPMFVINEMRRIARHLRFGNQEDAHE 200
Db 177 FHTSTDCRPTCMTCADDDMFSSYNSKNKSTFTFGPTAVLNMMLKLSKLCGYSQODGHE 236
QY 201 FLOQTVDMQKACLNGSNKLDRTQA---TTLVC-----QIFGGYLSRVKCLNCKGVSD 252
Db 237 FFVLLDQM-----HTESGGTSMPTCTPIHRIFSGSLKNVTVCLDCKERV 283
QY 253 TFDPIYDITLIEIKAAQSVNKALEOFVKPEQLDGENSYKSKCKMVPASKRFTIHRSSNV 312
Db 284 AVDPLMDISLDIN-EPTLQGLCLERFVSKEKY---QYSCSCSGSK-NAIKQLVDFDKLPPT 337
QY 313 LTLISLKEFA--NFT-GGKIAKDVVKPYEYLDIRPYMSQNGEPIVYVLYAVLVHTGNCNCHA 369
Db 338 ICMQLKRFQGNFAMSTKIDKQVSPAPLRWRYNFNQ---DDVDYQLYSVVCHKG-TLDT 393
QY 370 GHYFCYIKASNGLWYQNDISVSTDIRSVLSQOAVLVFIRSHDVKNKGGLTTPHSHS 427
Db 394 GHYIATYVQN-QWFLDLDITVIVEKSEVLNSQAYLLFY-HERQILYSDEMTVKTEN 449

RESULT 11

T46237
hypothetical protein T9C5.190 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46237
R;Rieger, M.; Gabel, C.; Mueller-Auer, S.; Schaefer, M.; Zipp, M.; Mewes, H.W.; Lemcke, F.
submitted to the Protein Sequence Database, December 1999
A;Reference number: Z23026
A;Accession: T46237
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1672 <RIE>
A;Cross-references: EMBL:AL132964
A;Experimental source: cultivar Columbia; BAC clone T9C5
C;Genetics:
A;Map position: 3
A;Introns: 591/1; 631/2; 829/2; 914/3; 1151/2; 1187/3; 1247/2; 1342/2; 1411/3; 1435/3; 15
A;Note: T9C5.190

Query Match 8.6%; Score 351.5; DB 2; Length 1672;
Best Local Similarity 23.7%; Pred. No. 8.4e-13;
Matches 129; Conservative 79; Mismatches 192; Indels 145; Gaps 21;
QY 3 IVDKASESSDFPSAYQNPQSGSEAVSPGDM DAGSAGSVSLSDYNSNHTLSLGPVGVAV 62
Db 595 VTGEISGGDTMSRPNTRKNKRQRPDAVDSSSQLIRKIHEANDVTDDDINQLFMWKPV 654

QY 63 YSSSSVPDKSPQKQDQALGDGIAPPKVLFPSEKICLKWQOTHR- 109
Db 655 CQCRVNRTRNPN-----CFCGLVPP---LNSRKGL-WQKISEIIQSLGDDPTLDRR 704
QY 110 -----GAGLQNLGNTCFANAALQCLTVTPPLANYMLSHRSKTKCHABGFCWMTQMAHIT 164
Db 705 DSESTPAGLNLGATCVANSILQCL-----YM-----NTAFREG-----VFSVEVHL 747
QY 165 QALSNPCDVLKPMFVINEMRIARHLRF-----VLDQIARLFAOLHASQKSFVDSDAFVKLTLELDNGVQODTHEFLTL 797
Db 748 K--QNP-----VLDQIARLFAOLHASQKSFVDSDAFVKLTLELDNGVQODTHEFLTL 797
QY 206 VDAQKACLNGSKNLDHRHTQATTLVCOIFGGYLSRVKCLNC---KGVSDTFDPYLDITL 262
Db 798 LSLERCLLHSGVK-----AKTIQDLFSGSVSHVTTCKCGORDSEASSKMDFALEL 851
QY 263 EIKAAQSVNKALQFOFKVPEQLDGENSYKCKCKKMWVPASKRFTIHRSSNVLTLSLKR--F 320
Db 852 NVKGLSLDASLNDYLSLEQLNGDNQVFCGSCNARVDATRCIKLRTLPPVITFQLKRCIF 911
QY 321 ANFTGG--KIADVKVPEYLDIRPYMSQPNGBPIVYVLYAVLVHTGFCHAGHYFCYIK- 377
Db 912 LPKTTAKKLTSSFPQVLDMSRLAESSQNKLTVDLSAVLHKGSVNSGHYVAHKD 971
QY 378 ASNGLWYQMNDSIVS-----TSDIRSVL-- 400
Db 972 EKTGLWFEFDEHVSGLKRPCKNEASSTPQBSNCTASSGNITDGIQSSSDCRSAIKS 1031
QY 401 -----SQAAVLFVIRSHDVKNVGELTHPTSPQ--SSRPVISQSVVTVNKAQAPGFIGPQ 455
Db 1032 EVFSSDAYMLYSLRCDKQENQ-----QKENPIDITKGEV---KQLKGGY---- 1076
QY 456 LPSHM 460
Db 1077 LPKHL 1081
RESULT 12
S63141
Hypothetical protein YNL186w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein N1619
C:Species: Saccharomyces cerevisiae
C>Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 19-Apr-2002
A:Accession: S63141
R:Obermaier, B.; Piravandi, E.; Rinke, M.; Dondey, H.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S63122
A:Accession: S63141
A:Molecule type: DNA
A:Residues: 1-792 <OBE>
A:Cross-references: EMBL:Z71462; NID:gl302172; PID:e239559; PID:gl302173; GSPDB:GN00014;
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:UBP10; MIPS:YNL186w
A:Cross-references: SGD:S0005130
A:Map position: 14L

Query Match 8.4%; Score 343.5; DB 2; Length 792;
Best Local Similarity 22.6%; Pred. No. 9.1e-13;
Matches 115; Conservative 78; Mismatches 174; Indels 141; Gaps 17;

QY 96 SEKICLKW--QOTHRAGLQNLGNTCFANAALQCLITYTPPLANYML-----SHRSKTC 148
Db 345 STRIVKMGDKFTNLKPRGLNHGVTCTYNAVAQMLHPSQHLVFDILMKGYDSTISK 404
QY 149 HAEGFCWMTQAHITQALSNP-----GOVFKPMFVINEMRIARHLRGNQEDAEHFLO 203
Db 405 NSVSTVLAETSKMWLPVSKNPRKNVSAIYNPKHLISRLDDINCWSEWQEDSHEVFM 464
QY 204 YTVDAQKACLNGSKNLDHRHTQATTLVCOIFGGYLSRVKCLNGKGVSDTFDPYLDITL 263
Db 465 SLMSRLQEDSVKPGHKL-----IESIYDIFGLLKQIVTCKSCGSISKTEQPFYDLSIH 519

QY 264 IKAQAQ----- 268
Db 520 LKQKKLDPNSDLSSDSINGTSATTSTTTTNAATKPSLSSSSVNLNMGSPFAAASDLSS 579
QY 269 -----SVNKALEQVVRPE--OLDGE--NSYKCKCKKMWVPASKRFTIHRSSNVLTLSLKR 320
Db 580 ANRRFIEKIKDFFPELIVKDKQKGYVCEKCHKTNAVKHSSILRAPETLLVHLKRF 639
QY 321 ANFTG--GKIADVKVPEYLDIRPYMSQPNGBPIVYVLYAVLVHTGFCHAGHYFCYIK 377
Db 640 -RENGTSSSKMKQAVSPMFLDLEY-CESKELPVKYLQLLSVVHGEGRSLSSGHIYAHCK 697
QY 378 ASNGLWYQMNDSIVSISRVSQ-QAYVLFVIRSHDVKNVGELTHPTSPQSSPRPV 436
Db 698 QPDGSWATYDDYDIIISERDVLKEPNAYLLYTR-----LT-----PKSVPLRL 742
QY 437 ISQVVTVNQAAPCFITGPQLPSHMIKNPPLHNGTGPLKOTPPSSSMSPNGSNSSVNRASPV 496
Db 743 AKSAMAT-----GNVTSKSKQEQAVNEPN-----NRPLKI 772
QY 497 NASA-SVQKNSVNFSSVIEPHEPKQKIT 523
Db 773 NSKKNRKKKKKN-----KKRKF 791

RESULT 13

B84611

probable ubiquitin carboxyl terminal hydrolase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
A:Accession: B84611
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; UID:20083487; PMID:10617197
A:Accession: B84611
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-365 <STO>
A:Cross-references: GB:AE002093; NID:g4567196; PIDN:AAD23612.1; GSPDB:GN00139

C:Genetics:
A:Gene: At2g22310
A:Map position: 2

Query Match 8.3%; Score 336; DB 2; Length 365;
Best Local Similarity 24.6%; Pred. No. 9e-13;
Matches 96; Conservative 61; Mismatches 135; Indels 98; Gaps 11;

QY 79 DOALGDGIAPPKVLFPSEKICLKWQOTHRAGLQNLGNTCFANAALQCLITYTPPLANY 138
Db 9 EKALGDGPPEGERYF-----GFENFGNTCYCNSVLQALYFCAPPREQ 50
QY 139 MLSHESKTKCHAEFCWMC--TMOAHIQALSNPCDVLKPMFVINEMRIARHLRGNQ 196
Db 51 LLEHYANKADAENLLTCLADLFSQISQKKKTG-VIAPKRFVORLKKONELFRSYMHQ 109
QY 197 DAHEFLOYTVDAQKACLNGSKNLDHRHTQAT----- 227
Db 110 DAHEFLNLLNLVEI-----LEKETQATKADNETSSSPEKIANVLKAPLANGVHKRP 162
QY 228 --TLVCOIFGGYLSRVKCLNGKGVSDTFDPYLDITLLEKAAQSVNKALQFOFKVPEQLDG 285
Db 163 IVTWVHKIFQGITNTRCLRCETVARDTDLSDLEIQNSITSLCKNFSSTETLHA 222
QY 286 ENSYKCKCKKMWVPASKRFTIHRSSNVLTLSLKRANFTG-----GKIADVKYP----- 335
Db 223 EDKFFCDKCCSLOEAQKMKIKPPHILVHLKRPKYMEQLGRYKCLSVRVFPLEKLS 282
QY 336 ----FYLDIRPYMSQPNGBPIVYVLYAVLVHTGFCHAGHYFCYIKASNGLWYQMN----- 388

Db 283 NTVDYVDIE-----YSLFAVVHVHVGSPNKGHGVSLVKSHNH-WLFFDDSEV 329

QY 389 STVSTDIRSVL-----SQQAYVLFY 409

Db 330 ELIESAVQTFPGSSQFVSSNTDGHYILLY 359

RESULT 14

S50277

ubiquitin-specific proteinase UBPS (EC 3.4.-.-) - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein YER144c

C:Species: Saccharomyces cerevisiae

C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 29-Sep-1999

C:Accession: S50277; S50647

R:Xiao, W.; Fontanie, T.; Tang, M.

Yeast 10, 1497-1502, 1994

A:Title: UBPS encodes a putative yeast ubiquitin-specific protease that is related to th

A:Reference number: S50277; PMID:7871889

A:Accession: S50277

A:Molecule type: DNA

A:Residues: 1-805 <XIA>

A:Cross-references: EMBL:U10082; NID:g5953376; PIDN:AAC48928.1; PID:g5953377

R:Dietrich, F.S.

submitted to the EMBL Data Library, December 1994

A:Description: The sequence of S. cerevisiae cosmids 8229, 9115, 9132, 9981, and lambda

A:Reference number: S50647

A:Accession: S50647

A:Molecule type: DNA

A:Residues: 1-805 <DIE>

A:Cross-references: EMBL:U18917; NID:g603377; PIDN:AAB64671.1; PID:g603384; MIPS:YER144c

C:Genetics:

A:Gene: SGD:UBP5

A:Cross-references: SGD:S0000946; MIPS:YER144c

A:Map position: 5R

C:Superfamily: deubiquinating enzyme SSV7

C:Keywords: hydrolase

Query Match 8.2%; Score 332; DB 2; Length 805;

Best Local Similarity 26.1%; Pred. No. 4.5e-12;

Matches 123; Conservative 75; Mismatches 162; Indels 112; Gaps 22;

QY 8 SESSDFSAVQNGPSSEAVSPGMDAGSAGVSLNVSNTHTLSLGPVPGVAVVSSSS 67

Db 376 SKLNTPTSTQNKANTVERISP-DIRAAQA-----HAY-----LPPASNVFSR 417

QY 68 VPKSKSPQKQALGDGIAPPQKVLFPSEKI CLKWQOQTHRVGAGLQNLGNTCFANAALQ 127

Db 418 IP-----PLPOONLS-----SSRQTLNNSQVLLD-----LIVGLEINIGCCVMNCILQ 462

QY 128 CLATVTPPLANYMLSHR-----SKTCHAEAGFCMCTMQAHTQALSNPGDVIKP 176

Db 463 CLVGTDLVRMFLDNTYLNFINFDSSRGSKGLLAKNFALLVN-NMHRGAFTPPNVRTIP 521

QY 177 MFVINEMRRIARHLR-----FGNQEDAHEFLQYTVDMQKAC-LNGSNKLDH----- 223

Db 522 VQTI-QFKKICGHINPMYSDMQDCQFCQFLDGLHEDLNQNGSKKHLQSLDEERM 580

QY 224 -----TQATTL-----VCOIFGGLRSRVKCLNCKGVSDTFDPLDITLIEIK 265

Db 581 REKMSIRKASALEWERLLTDFSAIIDLPQQYASRIQCQCEHTSTITYQFVSLVSPVP 640

QY 266 AAQSVN--KALEQFVKPQLDGENSYKSKCKKMPVASKRFTTHRSSNVLTSLSKRFANF 323

Db 641 RVKTCNILDCFREFTKCERLGVDEQWSCPKLKKQPSKQLKITRLPKKLIINLKRFDN- 699

QY 324 TGGKIAD---VKPEYLDIRPYMSQP-NGEPIV-----VLYAVLVHTGF 365

Db 700 ---QMNKNVQVPEYSLDLPYWARDFNHEIAVNEIDIPTRGQVPPPIRYLYGVACHSG- 755

QY 366 NCHAGHYFCYI-KASNGWYQNDSI-----VSTDIRSVLSQQAAYVLFYR 411

Db 756 SLYGGHYTSYVYKPKKGMWFFDDSLYRPITFT-----EFTPSAYVLFYER 803

RESULT 15

T06097

ubiquitin-specific proteinase (EC 3.4.-.-) UBPS - Arabidopsis thaliana

N:Alternate names: protein T5J17.80

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999

C:Accession: T06097

R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15184

A:Accession: T06097

A:Molecule type: DNA

A:Residues: 1-371 <BEV>

A:Cross-references: EMBL:AL035708; GSPDB:GN00062; ATSP:T5J17.80

A:Experimental source: cultivar Columbia; BAC clone T5J17

C:Genetics:

A:Gene: ATSP:T5J17.80; UBPS

A:Map position: 4

A:Introns: 39/3; 78/3; 111/3; 177/3; 238/2

C:Keywords: hydrolase

Query Match 7.9%; Score 323.5; DB 2; Length 371;

Best Local Similarity 25.4%; Pred. No. 5.1e-12;

Matches 99; Conservative 56; Mismatches 143; Indels 91; Gaps 12;

QY 79 DOALGDGIAPPQKVLFPSEKI CLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANY 138

Db 9 EKALGDQFPEGERYF-----GFENFGNTCYCNVQLQALYFCVPFREQ 50

QY 139 MLSH-----EHSKTCHAEAGFCMCTMQAHTQALSNPGDVIKPMFVINEMRIA 187

Db 51 LLEYTTSNKSVDABENLMTCLADLF-----SQISSQKKKTG-VIAPKRFVQRLKKQN 102

QY 188 RHLRFGNQEDAHEFLQY-----TVDMQKAC-----LNG-----SNKL 220

Db 103 ELFRSYMHQDAHEFLNVLNEVVDILLEKEAKATKTEHTSSSSSSPEKIANGLVQPOANGV 162

QY 221 DRHTQATTLVCOIFGGLRSRVKCLNCKGVSDTFDPLDITLIEIKAAQSVNKALEQFVKP 280

Db 163 VHKEPIVTVVHNIFQGIILTNETRLCETVTARDETFLDLSLDIEQNSSITSCILKNFSST 222

QY 281 EQLDGENSYKSKCKKMPVASKRFTTHRSSNVLTSLSKRFANFTG-----GKIADVKYPE 336

Db 223 ETLHAEDKFFCDKCCSLQEAQKMKIKPPIHLVIHLKRFKYIEQLGRYKLSYRVVFEPL 282

QY 337 YL-----DIRPYMSQPNGEPIVYLYAVLVHTGNCHAGHYFCYIKASNGWYQNDSI- 391

Db 283 ELKLSNTEVEPYAD-----VEYSLFAVVHVHVGSPNKGHGVSLVKSHNH-WLFFDDSEV 335

QY 392 -----STSDIRSVLSQQAAYVLFY 409

Db 336 MIEESAVQTFPGSSQFVSSNTDGHYILLY 364

Search completed: August 10, 2004, 20:00:19

Job time : 21 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:55:58 ; Search time 43 Seconds
(without alignments)
5686.665 Million cell updates/sec

Title: US-10-049-745-4

Perfect score: 4070

Sequence: 1 MTIVDKASESSDPSAYQNOP.....AAESLEEDPAAASLRPFSEG 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1808.5	44.4	1016	13 Q802X0	Q802X0 brachydanio
2	1375	33.8	548	4 Q9H9C5	Q9H9C5 homo sapien
3	915	22.5	919	5 Q86JL6	Q86JL6 dictyosteli
4	895	22.0	466	11 Q8BTN5	Q8BTN5 mus musculu
5	894	22.0	545	11 Q923V2	Q923V2 mus musculu
6	890.5	21.9	545	11 Q923V3	Q923V3 mus musculu
7	889.5	21.9	545	11 Q55191	Q55191 mus musculu
8	888	21.8	545	11 Q55190	Q55190 mus musculu
9	763.5	18.8	842	10 Q84P80	Q84P80 mus musculu
10	760	18.7	457	3 Q74442	Q74442 oryza sativ
11	748	18.4	859	10 Q9FPS4	Q9FPS4 schizosacch
12	737	18.1	661	10 Q9FPS2	Q9FPS2 arabidopsis
13	735.5	18.1	1038	5 Q8IQ60	Q8IQ60 drosophila
14	735.5	18.1	1085	5 Q9VRP5	Q9VRP5 drosophila
15	734	18.0	661	10 Q8VZF5	Q8VZF5 arabidopsis
16	733.5	18.0	948	5 Q8IQ58	Q8IQ58 drosophila

17	733.5	18.0	968	5 Q96Q04	Q96Q04 drosophila
18	733.5	18.0	995	5 Q8IQ59	Q8IQ59 drosophila
19	733.5	18.0	1015	5 Q8IQ57	Q8IQ57 drosophila
20	730.5	17.9	1038	5 Q86NM9	Q86NM9 drosophila
21	725.5	17.8	948	5 Q7YTX7	Q7YTX7 drosophila
22	720	17.7	891	10 Q9SHG9	Q9SHG9 arabidopsis
23	704	17.3	672	10 Q9SJA1	Q9SJA1 arabidopsis
24	692.5	17.0	631	10 Q81780	Q81780 arabidopsis
25	686.5	16.9	731	10 Q9FKP5	Q9FKP5 arabidopsis
26	686.5	16.9	924	10 Q9FPS9	Q9FPS9 arabidopsis
27	646.5	15.9	632	10 Q8LUK9	Q8LUK9 arabidopsis
28	645	15.8	829	10 Q9FUL7	Q9FUL7 arabidopsis
29	626.5	15.4	1008	10 Q9FPS8	Q9FPS8 arabidopsis
30	626.5	15.4	1008	10 Q9SB51	Q9SB51 arabidopsis
31	555.5	13.6	285	4 Q8IXW9	Q8IXW9 homo sapien
32	555.5	13.6	732	10 Q9FIQ1	Q9FIQ1 arabidopsis
33	555.5	13.6	737	10 Q9FPS6	Q9FPS6 arabidopsis
34	504.5	12.4	695	10 Q9FPS7	Q9FPS7 arabidopsis
35	504.5	12.4	1082	10 Q49688	Q49688 arabidopsis
36	437	10.7	703	5 Q95TK9	Q95TK9 drosophila
37	437	10.7	735	5 Q9VUQ9	Q9VUQ9 drosophila
38	437	10.7	735	5 Q9VVR1	Q9VVR1 drosophila
39	420	10.3	520	4 Q8WVD0	Q8WVD0 homo sapien
40	415.5	10.2	525	11 Q9JIG5	Q9JIG5 mus musculu
41	413	10.1	938	5 Q8IQ27	Q8IQ27 drosophila
42	411	10.1	856	5 Q9VR54	Q9VR54 drosophila
43	403	9.9	337	5 Q8IQ26	Q8IQ26 drosophila
44	402.5	9.9	540	11 Q8BSW2	Q8BSW2 mus musculu
45	393	9.7	896	5 Q8MQX4	Q8MQX4 drosophila

ALIGNMENTS

RESULT 1

Q802X0 PRELIMINARY; PRT; 1016 AA.

AC Q802X0; DT 01-JUN-2003 (TREMREL. 24, Created)

DT 01-JUN-2003 (TREMREL. 24, Last sequence update)

DT 01-OCT-2003 (TREMREL. 25, Last annotation update)

DE Similar to KIAA1453 protein (Fragment).

OS Brachydanio rerio (Zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

OX NCBI_TaxID=7955;

[1] SEQUENCE FROM N.A.

RP STRAIN=AB;

RC Strausberg R.;

RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC047168; AAH47168.1; -

DR GO; GO:0004197; P:cysteine-type endopeptidase activity; IEA.

DR GO; GO:0004379; P:glycylpeptide N-tetradecanoyltransferase activity; IEA.

DR GO; GO:0004221; P:ubiquitin thioesterase activity; IEA.

DR GO; GO:0006499; P:N-terminal protein myristoylation; IEA.

DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.

DR InterPro; IPR003006; Iq MHC.

DR InterPro; IPR000903; Nmt.

DR InterPro; IPR001394; Peptidase_C19.

DR Pfam; PF00443; UCH; 1.

DR PROSITE; PS00290; IG MHC; 1.

DR PROSITE; PS00976; NMT 2; 1.

DR PROSITE; PS00972; UCH_2_1; 1.

DR PROSITE; PS00973; UCH_2_2; 1.

DR PROSITE; PS00973; UCH_2_3; 1.

FT NON TER 1016 1016

SQ SEQUENCE 1016 AA; 113138 MW; 05AFD86206CC9DAA CRC64;

Query Match 44.4%; Score 1808.5; DB 13; Length 1016;

Best Local Similarity 50.9%; Pred. No. 6.3e-120;

Matches 390; Conservative 96; Mismatches 199; Indels 81; Gaps 20;

RT "NEDO human cDNA sequencing project";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK022913; BAB14306.1; -
 DR GO; GO:0004197; F:cytisine-type endopeptidase activity; IEA.
 DR GO; GO:0004221; F:ubiquitin thioesterase activity; IEA.
 DR GO; GO:0006511; F:ubiquitin-dependent protein catabolism; IEA.
 DR InterPro; IPR003006; Ig MHC.
 DR InterPro; IPR001394; Peptidase_C19.
 DR Pfam; PF00443; UCH; 1.
 DR PROSITE; PS00290; IG_MHC; 1.
 DR PROSITE; PS00972; UCH 2.1; 1.
 DR PROSITE; PS00973; UCH 2.2; 1.
 DR PROSITE; PS00235; UCH 2.3; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 548 AA; 60332 MW; 2CSED382105AC2C6 CRC64;
 Query Match 33.8%; Score 1375; DB 4; Length 548;
 Best Local Similarity 47.3%; Pred. No. 1.9e-89;
 Matches 286; Conservative 81; Mismatches 152; Indels 86; Gaps 12;
 QY 1 MTIVKASSESDPEAYONOPGSEAVSPGDMAGSASGAVSSLDV----- 47
 DB 1 MPVLDKLEAL-----KPERKDSADGELGKLLASSAKVLLQKIEFEPASKSFYQL 53
 QY 48 ----SNTLSLGPVPGAVVY-SSSVDPKSPQKQDQALGDIAPQKVLFPSEKICLK 102
 DB 54 EALKSKYVLLNPKTEGASRHKSGDDPPARRRGSEHTYESCGDGVPAQKVLFPETERLSLR 113
 QY 103 WQOQTHRVGAGLONIGNTCFANALQCLTYTPPLANYMLSHSEKTCCHAEFGCMCTMQAH 162
 DB 114 WERVFRVAGLHNLGNTCFNLATIQCLTYTPPLANYLLSEHARSCHQGSFCLMCMQNH 173
 QY 163 ITQALSNPGDVKPMFVINEMRRIARHLRFGNEDAEHEFLQYTVDMOKACLNGSNKLD 222
 DB 174 IVQAPNSGNAIKPVSFIRDLKLIARHFRFGNEDAEHEFLRYTIDAMOKACLNGCAKLD 233
 QY 223 HTQATTLVCOIFGYSRLSRVKCLNGKGVSDTFDPYDITLILKAAQSVNKALEOFVKPEQ 282
 DB 234 QTQATTLVHQIFGYSRLSRVKCSVSDTYDPLDVALEIRQAAIVRALELFRKADV 293
 QY 283 LDGENSKCSKCKQWPAKRFTHRSNNVTLTLKRFANFTGGKIAKDVKYPEYLDIRP 342
 DB 294 LSGENAYMCAKCKKVPASKRFTIHRNSVTLTLKRFANFTGGKIAKDVKYPEYLDIRP 353
 QY 343 YMSQNGEPIVVLVAVLVHTGFNCHAGHYFCYIKASNGLWYQNDISVTSDFSRVLSQ 402
 DB 354 YMSQNGDPVMYGLVAVLVHSGYCHAGHYCYVKASNGQWYQNDISVHSSNVKVLNQ 413
 QY 403 QAYVLYIRSHDVKNKGELTHPTSPG-QSSPRPVISORVVTNKQAARPGTGPQ-LPSHM 460
 DB 414 QAYVLYLR-----IPGSKSPGELISR---TGSSELPF--RPSVIPDHS 453
 QY 461 IKNPPLNGTGPL--KOTPSSSMSSPNSGNSVNRASPVNASASVQNSWNSRVSVIPEHPK 518
 DB 454 KKNIGNGIISPLTKKQKQDSGTMKKPHTEI--GVPIRNGSTLGLKSQNGCIPKPLPS 511
 QY 519 KOKITISHNKLPVRQCSQPNLHNSLENTPKVPSSITNSAVQSTNSASTMVSVKY 578
 DB 512 GS-----FSPKLSQTPTHTMPTILDDPK-----KV 536
 QY 579 TKPIP 583
 DB 537 KKPAF 541
 RESULT 3
 Q86J16 PRELIMINARY; PRT; 919 AA.
 ID Q86J16
 AC Q86J16; (TREMELrel. 24, Created)
 DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
 DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Hypothetical protein.

QY 31 MDAGSASGAVSSLDVSN-HTLSLGPVPGAVVYSSSV-DKSPKSPQKQALGDGTAP 88
 DB 1 MESSSSSVGSAFELSRAKTVCMAPALGATVNGSNNTALQVERPREQVGTSGDIAL 60
 QY 89 POKVLFPSEKICLKWQOQTHRVGAGLONIGNTCFANALQCLTYTPPLANYMLSHSEKTC 148
 DB 61 POKVLFPSEKICLKWQOQTHRVGAGLONIGNTCFANALQCLTYTPPLANYMLSHSEKTC 120
 QY 149 HAEGCMCTMQAHTLQALSNPGDVKPMFVINEMRRIARHLRFGNEDAEHEFLQYTVDA 208
 DB 121 HEPGFCMCTMQAHTLQALSNPGDVKPMFVINEMRRIARHLRFGNEDAEHEFLQYTVDA 180
 QY 209 MQKACLNGSNKLDHRTQATTLVCOIFGYSRLSRVKCLNGKGVSDTFDPYDITLILKAAQ 268
 DB 181 MQKACLNGSNKLDHRTQATTLVCOIFGYSRLSRVKCLNGKGVSDTFDPYDITLILKAAQ 239
 QY 269 SVNKALRQVKEPQDGENSYKSKCKQWPAKRFTHRSNNVTLTLKRFANFTGGKI 328
 DB 240 TLSKAPEQFVKPQLDGDNAKCKSKCKQWPAKRFTHRSNNVTLTLKRFANFTGGKI 299
 QY 329 AKDVKPEYLDIRPYMSQNGEPIVVLVAVLVHTGFNCHAGHYFCYIKASNGLWYQND 388
 DB 300 TKDVRVTAELDLRPFMSQSGEPQIYALVAVLVHSGFCHAGHYCYVKASNGQWYQND 359
 QY 389 SIYSTDIRSVLSQAYVLYIRSHDVKNKGELTHPTSPGQSSPRPVISORVVTNKQAA 448
 DB 360 SSVLSDIRVLSQAYVLYIRSHDVKNKGELTHPTSPGQSSPRPVISORVVTNKQAA 419
 QY 449 PGFIPQLPSHMLKNPPLNGTGPLKDTPSSMSNGN-SSVNRASPVNASVQWQSV 507
 DB 420 TSFIPQLPSHMLKNPPLNGTGPLKDTPSSMSNGN-SSVNRASPVNASVQWQSV 477
 QY 508 N---RSSVIEHPKPKKITTISHNKLPV--QCOSQPNLHNSLENTPKVPSSITNSA 562
 DB 478 SLVRSWGIIPDSKRPKLTFLIGOGKVPRAQTSQSPNCSLSSA-----SH 523
 QY 563 VQSTSNASTMSVSKVTKPIPRSSCSQPVNKGSKLNSVLYVYGAESSEDSDESKGL 622
 DB 524 SQSTSSSTSTSG-----PQVQKQ-----VNG-THTGASFLVYGOESSEDSDESKGL 572
 QY 623 GKENG-----IGTIVSHSPQDAE--DEATPHELOEPMPLANGANSAD--SDSDPEN 672
 DB 573 --ENGSTKTYLAPKITNGNGVYKDDAQSHSSSLPHHTSKTNGSGNHFVNGSGSHVCEN 630
 QY 673 GLAPDGASQCOQALHNSNPFKANGLPGLKMPAPLPSLPEDKILETFRLSNKLKG----- 728
 DB 631 GSGPDHT-----SQNCLSKANGFN-----HNDKVMIDPHTSSQISDLSCC 670
 QY 729 STDMSAPGAERGPPEDRDAPQP-----GSPAASLEEDPAAS 768
 DB 671 SEMELNKPANH---ESRSSNPASSDDHSCSPSNKRLNLPSESS 713
 RESULT 2
 Q9H9C5 PRELIMINARY; PRT; 548 AA.
 ID Q9H9C5
 AC Q9H9C5; (TREMELrel. 16, Created)
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Hypothetical protein FLJ12851.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Isoqai T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 RA Tanase T., Nomura Y., Togiya R., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.,

Q923V2		PRELIMINARY;	PRT;	545 AA.
ID	Q923V2;			
AC	Q923V2;			
DT	01-DEC-2001 (TREMBlrel. 19, Created)			
DT	01-DEC-2001 (TREMBlrel. 19, Last sequence update)			
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)			
DE	Deubiquitinating enzyme 2A.			
GN	DUB2 OR DUB2A.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129/W;			
RA	Baek K.-H., Mondoux M.A., Jaster R., Fire-Levin E., D'Andrea A.D.;			
RT	"DUB-2A, a new member of the DUB subfamily of hematopoietic			
RT	deubiquitinating enzymes."			
RL	Blood 98:0-0(2001).			
DR	EMBL; AF393638; AAK77003.1; --.			
DR	MGD; MGI:107698; Dub2.			
DR	GO; GO:0004843; F:ubiquitin-specific protease activity; IDA.			
DR	GO; GO:0016579; P:protein deubiquitination; IDA.			
DR	InterPro; IPR001394; Peptidase_C19.			
DR	Pfam; PF00443; UCH; 1.			
DR	PROSITE; PS00972; UCH_2_1; 1.			
DR	PROSITE; PS00973; UCH_2_2; 1.			
DR	PROSITE; PS0235; UCH_2_3; 1.			
SQ	SEQUENCE 545 AA; 2FFDF097313FB88D CRC64;			
Query Match 22.0%; Score 894; DB 11; Length 545;				
Best Local Similarity 36.3%; Pred. No. 3.3e-55;				
Matches 208; Conservative 96; Mismatches 193; Indels 76; Gaps 14;				
QY	69	PDKSPSPQ---KDQALGDGIAPPQKV--LFPESEKICLKWQTHRVGAGLQNLGNTCFAN	123	
DB	12	PALSSPGAQLHQDEA-----QVVVELTANDKPSLSWECPCPGCGLQNTGNSCYLN	63	
QY	124	AAQLCYTPPLANMYLSHEHSTKHAEGFCMMCTQAHITQAL--SNPGDVIKPMFVIN	181	
DB	64	AAQLCYTHPTPLADYMLSQEYSQTCSPGCKKAMEAHVTSLLHSHSGDVNPKSQILT	123	
QY	182	EMRRIARHLRFNGOEDAHFLOQTVDMOKACILNGSKLDRHTQATTLVCQIFGGYLSR	241	
DB	124	S-----AFHXH--QOEDAHEFLMFTLETMHESCLQVHRQSEPTSEDSPPHDI	177	
QY	242	VKCLNCKGVSDTPDPYLDITLTKAAQSVNKALEQFVKPEQLDGENSYKSKCKKMPAS	301	
DB	178	IKLHCQGTSDYDRFLDPLDISSAQSVNQALWDTEKSEELGENAYYCGRCQKMPAS	237	
QY	302	KRFTIHRSSNVLTSLKRFANFTGGKIADVKYPEYLDIRPYMSQNGEPIVIVLYAVLV	361	
DB	238	KTLLHISAPKVLVLLKRFSAFNGKLDKRVSPYEFDLKPYLSQPTGGPLPYALYAVLV	297	
QY	362	HTGFNCHAGYFCYIKASNGLYQWMDNSIVSTDIRSVLSQAAVLYFIIRSHDVXNGGEL	421	
DB	298	HEGATCHSGHYFSYVAGHGKWKMDTKVTSCDVTSLNENAYLVFYVQOTDLKE-----	353	
QY	422	THPTSPGQSSPRPVISQRVVTNKAAPGFIGLPQPS---HMTKNPPH--LNGTGPKLQTP	477	
DB	354	-----VSDMPEGRHIEVLDPYQLKRRKKHKK	383	
QY	478	SSMSPPNGNSVNRASPVNASASVQWNSVNRSSVIEPHKQKQITISIHKLTVRCQS	537	
DB	384	KSPCTEDVGSPKNR-----EKKATKETSLEGKVLQEKHKKAGQKHNTKL--VPEQON	437	
QY	538	QPNLHNSLNPPTKPVPS--TITNSAVQSTNSASTMSVSKVTPIPRS-----ES	587	
DB	438	HQKLQGHKRNELLPOENHQKQKQSLNTEGELDLPADAIVH--LFRSIANWGRDTPDK	496	
QY	588	CSQPVNMGKSLNSSLVLPYGAESSEDSDESK	620	
DB	497	VNQFWHNADRLILTSDQLVNTGLLCKQEGRRRSK	529	

Q923V3		PRELIMINARY;	PRT;	545 AA.
ID	Q923V3;			
AC	Q923V3;			
DT	01-DEC-2001 (TREMBlrel. 19, Created)			
DT	01-DEC-2001 (TREMBlrel. 19, Last sequence update)			
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)			
DE	Deubiquitinating enzyme 2A.			
GN	DUB2 OR DUB2A.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129;			
RA	Baek K.-H., Mondoux M.A., Jaster R., Fire-Levin E., D'Andrea A.D.;			
RT	"DUB-2A, a new member of the DUB subfamily of hematopoietic			
RT	deubiquitinating enzymes."			
RL	Blood 98:636-642(2001).			
DR	EMBL; AF393637; AAK84135.1; --.			
DR	MGD; MGI:107698; Dub2.			
DR	GO; GO:0004843; F:ubiquitin-specific protease activity; IDA.			
DR	GO; GO:0016579; P:protein deubiquitination; IDA.			
DR	InterPro; IPR001394; Peptidase_C19.			
DR	Pfam; PF00443; UCH; 1.			
DR	PROSITE; PS00972; UCH_2_1; 1.			
DR	PROSITE; PS00973; UCH_2_2; 1.			
DR	PROSITE; PS0235; UCH_2_3; 1.			
SQ	SEQUENCE 545 AA; 61339 MW; 0154DF343C8DFIA5 CRC64;			
Query Match 21.9%; Score 890.5; DB 11; Length 545;				
Best Local Similarity 35.5%; Pred. No. 5.8e-55;				
Matches 210; Conservative 91; Mismatches 191; Indels 99; Gaps 14;				
QY	69	PDKSPSPQ---KIQALGDGIAPPQKV--LFPESEKICLKWQTHRVGAGLQNLGNTCFAN	123	
DB	12	PALSSPGAQLHQLEA-----QVVVELTANDKPSLSWECPCPGCGLQNTGNSCYLN	63	
QY	124	AAQLCYTPPLANMYLSHEHSTKHAEGFCMMCTQAHITQAL--SNPGDVIKPMFVIN	181	
DB	64	AAQLCYTHPTPLADYMLSQEYSQTCSPGCKKAMEAHVTSLLHSHSGDVNPKSQILT	123	
QY	182	EMRRIARHLRFNGOEDAHFLOQTVDMOKACILNGSKLDRHTQATTLVCQIFGGYLSR	241	
DB	124	S-----AFHXH--QOEDAHEFLMFTLETMHESCLQVHRQSEPTSEDSPPHDI	177	
QY	242	VKCLNCKGVSDTPDPYLDITLTKAAQSVNKALEQFVKPEQLDGENSYKSKCKKMPAS	301	
DB	178	IKLHCQGTSDYDRFLDPLDISSAQSVNQALWDTEKSEELGENAYYCGRCQKMPAS	237	
QY	302	KRFTIHRSSNVLTSLKRFANFTGGKIADVKYPEYLDIRPYMSQNGEPIVIVLYAVLV	361	
DB	238	KTLLHISAPKVLVLLKRFSAFNGKLDKRVSPYEFDLKPYLSQPTGGPLPYALYAVLV	297	
QY	362	HTGFNCHAGYFCYIKASNGLYQWMDNSIVSTDIRSVLSQAAVLYFIIRSHDVXNGGEL	421	
DB	298	HEGATCHSGHYFSYVAGHGKWKMDTKVTSCDVTSLNENAYLVFYVQOTDLKE-----	353	
QY	422	THPTSPGQSSPRPVISQRVVTNKAAPGFIGLPQPSHMTKNPPH--LNGTGPKLQTPSS	480	
DB	354	-----CSIDMPEGRHIEVLDPYQLKRRKKHKKSP	386	
QY	481	MSPNGNSVNRASPVNASASVQWNSV-----NRSSVIEPHKQKQITISIH	527	
DB	387	CTEDVGSPKRNREIKATKETSLEGKVLQEKHKKAGQKHNTKLVPQENHOKLQKH	446	
QY	528	NK--LPVRCQSQ?NLHNSLNPPTKVPSPSTTNSAVQSTNSASTMSVSKVTPIPRS	585	
DB	447	NNEILPQEQ-----NHQKTSQSLNTEGELDSPADA-----IVHLP	485	

```
QY 586 -----BSCQPMNGKSLNSSLVLPYGAESSEDSDEES-KGLGKEN 626
Db 486 IANWGRDTPDKVNPWHNADRLLTSQDLVNTGLLCROEGRRRSKGKNKN 536

RESULT 7
O55191 PRELIMINARY; PRT; 545 AA.
AC O55191
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hematopoietic-specific IL-2 deubiquitinating enzyme.
GN DUB2 OR DUB-2
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97150690; PubMed=8995226;
RA Zhu Y., Lambert K., Corless C., Copeland N.G., Gilbert D.J.,
RA Jenkins N.A., D'Andrea A.D.;
RT "DUB-2 is a member of a novel family of cytokine-inducible
RT deubiquitinating enzymes.";
RL J. Biol. Chem. 272:51-57(1997).
DR EMBL; U70369; AAB94636.1; -.
DR MGD; MGI:107698; DUB2.
DR GO; GO:0004843; F:ubiquitin-specific protease activity; IDA.
DR GO; GO:0016579; P:protein deubiquitination; IDA.
DR InterPro; IPR001394; Peptidase_C19.
DR Pfam; PF00443; UCH_1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS0235; UCH_2_3; 1.
DR PROSITE; PS0235; UCH_2_3; 1.
SQ SEQUENCE 545 AA; 61537 MW; 4FFDCDA7F182E206 CRC64;

Query Match 21.9%; Score 889.5; DB 11; Length 545;
Best Local Similarity 35.4%; Pred. No. 6.9e-55;
Matches 205; Conservative 93; Mismatches 190; Indels 91; Gaps 14;

QY 76 PQKDQALGDGIAPPQKVLPFSE-----KICLKWOQTHRVGAGLQNLGNTCFANA 124
Db 8 PEQDPALS-----SFGAQLHDEAQQVVVELTANDKPSLSWECPPQCGCLQNTGNSCYLNA 64
QY 125 ALQCLFTPTPLANYMLSHHSKTCBAGFCMCTMQAHITQAL--SNPGDVIRKPMFVINE 182
Db 65 ALQCLTHTPTPLADYMLSQEYSQTCSPGCKMCAMEAHVTQSLHSHSGDVWKPQSILTS 124
QY 183 MRRIARHLRFGNQEDAHEFLQYTVDMQKACLNGSNKLDRTQATTLVCOIFGGYLSRV 242
Db 125 ----AFKH--QOEDAHEFLMFTLETMHESCLQVHRQSEPTSDSSPIHDFIGGLWRSQI 178
QY 243 KCLNCKGVSDFPTPYLDITLLEIKAAQSVNKALEQVFKPEQLDGENSYKCSCKKQWVPASK 302
Db 179 KCLHCOGTSDFRDLVDPLDISASQSVNQALWDTEKSELGENAYTCRCQKQWVPASK 238
QY 303 RFTIHRSSNVLTLSLKRANFTGGKIADKVPEYLDIRPYMQNPGEPIVYVLYAVLVH 362
Db 239 TLHIHSAKPVLLVLRFSAFMGKLDKRVSYPEFLDKPYLSOPTGGPLPYALYAVLVH 298
QY 363 TGFNCHAGHYFCYIKASNLGYQMNDSIVSTSDIRSVLSQAAYLVFYIRSHDVNGGELT 422
Db 299 EGATCSGHYFSYVKARHGAWYKMDTKVTSQDVTSVLNENAYLVFYVQQTDLKQ----- 353
QY 423 HPTHSPGQSSPRPVISQSVVTKQAAPGFTGPOLPS---HMKNPPH-LNGTGPLKDTPS 478
Db 354 -----VSIDMPGVRVHEVLDPEYQLKSRKKHKKK 384
QY 479 SSMSPNGNSVNRASFPVNASVQNMV-----NRSSVIPHPKKQKI--T 523
Db 385 SPCTEDAGPCNKRKATKETSILGEGKVQENKHKAGQKHTKLVPOEQNHQKLGQK 444
```

```
QY 524 ISIHKLPLVRQCSQPNLHNSLENP--TKVPSPSTITNSAVQSTSNASTMSVSKVTKP 581
Db 445 HRINEILPQEQNHOKA---GQSLFNTGELDLPADAIVLHLLRSTENWG-----RD 492

QY 582 IPRESQSPQWNGKSLNSSLVLPYGAESSEDSDEESK 620
Db 493 APDKE--NQPMHNADRLLTSQDPVNTGOLCRQGRRRSK 529

RESULT 8
O55190 PRELIMINARY; PRT; 545 AA.
AC O55190
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hematopoietic-specific IL-2 deubiquitinating enzyme.
GN DUB2 OR DUB-2
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=129/Sv;
RX MEDLINE=97150690; PubMed=8995226;
RA Zhu Y., Lambert K., Corless C., Copeland N.G., Gilbert D.J.,
RA Jenkins N.A., D'Andrea A.D.;
RT "DUB-2 is a member of a novel family of cytokine-inducible
RT deubiquitinating enzymes.";
RL J. Biol. Chem. 272:51-57(1997).
DR EMBL; U70368; AAB95194.1; -.
DR MGD; MGI:107698; DUB2.
DR GO; GO:0004843; F:ubiquitin-specific protease activity; IDA.
DR GO; GO:0016579; P:protein deubiquitination; IDA.
DR InterPro; IPR001394; Peptidase_C19.
DR Pfam; PF00443; UCH_1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS0235; UCH_2_3; 1.
DR PROSITE; PS0235; UCH_2_3; 1.
SQ SEQUENCE 545 AA; 61480 MW; F2F01AB7F182E203 CRC64;

Query Match 21.8%; Score 888; DB 11; Length 545;
Best Local Similarity 35.7%; Pred. No. 8.8e-55;
Matches 207; Conservative 93; Mismatches 190; Indels 90; Gaps 15;

QY 69 PDKSKPSQ---KQDALGDGIAPPQKV--LFPSEKICLKWOQTHRVGAGLQNLGNTCFAN 123
Db 12 PALSSFGAQLHDEA-----QVVVELTANDKPSLSWECPPQCGCLQNTGNSCYLN 63
QY 124 ALQCLFTPTPLANYMLSHHSKTCBAGFCMCTMQAHITQAL--SNPGDVIRKPMFVIN 181
Db 64 ALQCLTHTPTPLADYMLSQEYSQTCSPGCKMCAMEAHVTQSLHSHSGDVWKPQSILTS 123
QY 182 EMRIARHLRFGNQEDAHEFLQYTVDMQKACLNGSNKLDRTQATTLVCOIFGGYLSRV 241
Db 124 S----AFKH--QOEDAHEFLMFTLETMHESCLQVHRQSEPTSDSSPIHDFIGGLWRSQ 177
QY 242 VKLNCCKGVSDFPTPYLDITLLEIKAAQSVNKALEQVFKPEQLDGENSYKCSCKKQWVPAS 301
Db 178 IKCLHCOGTSDFRDLVDPLDISASQSVNQALWDTEKSELGENAYTCRCQKQWVPAS 237
QY 302 RFTIHRSSNVLTLSLKRANFTGGKIADKVPEYLDIRPYMQNPGEPIVYVLYAVLVH 361
Db 238 TLHIHSAKPVLLVLRFSAFMGKLDKRVSYPEFLDKPYLSOPTGGPLPYALYAVLVH 297
QY 362 HTGFNCHAGHYFCYIKASNLGYQMNDSIVSTSDIRSVLSQAAYLVFYIRSHDVNGGEL 421
Db 298 HEGATCSGHYFSYVKARHGAWYKMDTKVTSQDVTSVLNENAYLVFYVQQTDLKQ---- 353
QY 422 TPTHSPGQSSPRPVISQSVVTKQAAPGFTGPOLPS---HMKNPPH-LNGTGPLKDTP 477
```



```

QY 131 YTPPLANTYMLSHESKTCBAEGFCMMCTMQAHITQALSNPG---DVIKPMFVINEMERRI 186
Db 145 QTPPLVQLYLSQHSLSCEMNA-CVLCMEQHVARYPNKGTKRASAFKPSGIQSMKVI 203
QY 187 ARHLRFGNQEDAHEFLQYTVDMQKACLNKGNKLDHRHTQATLVQCIFFGYLRSRVKCLN 246
Db 204 SSHFRPYRQEDAHEFMYLVDMAWQKSLQNHKNLDHPDRETSSVHRIFGGYLRQOILCSV 263
QY 247 CKGVSDTFDPVLDITLEIKAAQSVNKALEQFVKPQDGENSYKSKCKKMWVPASKRFTI 306
Db 264 CKKPSNTYQALLDLSVDAGKS-SLADSLKHFVHAEKLTQNKYRCENCQQLVDASKQMTI 322
QY 307 HRSSNVLTSLKREA--NFTGGKIAKDVKYPEYLDIRPYMSQNGEPIVYVLYAVLVHTG 364
Db 323 YRAPNLTIFKRTFNGCFQSKISKQISYPESFNLGPMYDPMCS-CWYELIGVLVHAG 381
QY 365 FNCHAGHYFCVIKASGLWYQNDISVSTDIRSVLSQQAAYVLFYIR 411
Db 382 GSTRSGHYSPCKSSNGWLKFDFFVSNSSIDRVLNQAYILQYKR 428

RESULT 11
Q9FPS4
ID Q9FPS4 PRELIMINARY; PRT; 859 AA.
AC Q9FPS4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
DE Ubiquitin-specific protease 23.
GN UBP23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20567829; PubMed=11115897;
RA Van N., Doelling J.H., Falbel T.G., Durski A.M., Vierstra R.D.;
RT "The Ubiquitin-Specific Protease Family from Arabidopsis. AtUBP1 and 2
Are Required for the Resistance to the Amino Acid Analog Canavanine.";
RL Plant Physiol. 124:1828-1843(2000).
DR EMBL; AF302671; AAG42761.1; -.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004221; P:ubiquitin thiolesterase activity; IEA.
DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
DR Pfam; PF00443; UCH_1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS00235; UCH_2_3; 1.
DR Protease.
KW Protease.
SQ SEQUENCE. 859 AA; 94889 MW; 4CC8E082B54CFE9 CRC64;

Query Match
Best Local Similarity 18.4%; Score 748; DB 10; Length 859;
Matches 201; Conservative 125; Mismatches 272; Indels 88; Gaps 19;

QY 67 SVDPKSPKQDQALGDGIAPQKVLFPSEKICLKWQTHR-VGAGLQNLGNTCFANAA 125
Db 75 SAFSANK--PDSSDLLEHGFED-----LTFSTFRKIGAGLQNLGNTCFNSV 121
QY 126 LQCLTYTPPLANTYMLSHESKTCBAEGFCMMCTMQAHITQALSNPGDVIKPMFVINEMRR 185
Db 122 LQCLTYTEPLAATLQTAHQKCHVAGFCALCAIQXHVRTARQANGRIILKDLVSNLRC 181
QY 186 IARHLRFGNQEDAHEFLQYTVDMQKACIL-NGSNKLDHRHTQATLVQCIFFGYLRSRVK 244
Db 182 ISRNFRNCRQEDAHEYMINLLECMKCSLPSGVPSSESDAYRRSLVHKIFGGLRSQVKC 241
QY 245 LNCKGVSDFDPVLDITLEIKAAQSVNKALEQFVKPEQLD-GENSYKSKCKKMWVPASKR 303

```

```

Db 242 EOCSCSNKFPFDLSDISKAQSLQALGRFTAVELLDNGAKVYQCERCKQKAKKQ 301
QY 304 FTIHRSSNVLTSLKRPANFTGGKIAKDVKYPEYLDIRPYMSQNGEPIVYVLYAVLVHT 363
Db 302 LTVSKAPVVLVHLKRFEAHSEKIDRVDTSAIDMKPFVSGPHEGMLKYLYIGVLVHY 361
QY 364 GFNCHAGHYFCVIKASGLWYQNDISVSTDIRSVLSQQAAYVLFYIRSHDVKNKGELTH 423
Db 362 GRSSHSGHYACFVRTSSGMWYSLDDNRVVQVSEKTVFNQKAYMLFYVR--DRQNA----- 414
QY 424 PTHSGGSSPPVLSQRVVTKNQAPGFIGQLPDSHMTKPNPHLNGTGLKDTPESSMS 483
Db 415 ---VPMKSV--VVKKESFATNRSL-----IVASNIKD--QVNGSTVIKCGFGALVA 461
QY 484 PNGNSSVNRASPVNASASVQ---NWSVNRSSVIEHPKKOKITISIHKNLPRVROQ---- 536
Db 462 -NGLAPLKSCGSPTPAVLTQKDLNAKETQNNAI--SNVEAKEILLETENGSAVPKTCDLAAP 519
QY 537 ---SQPNLHNSLENPTKVPVPSSTITNSAVOSTSNASTMSVSSKVTKEPIRSESCSQPM 593
Db 520 TVLVQKDLNTKEIFQKEVPLPQANGEGSLVKEDSKAACILPEKVS-----PHL 568
QY 594 NGKSKLNSVLVPYGAESSEDSDEESKGLKENG-----IGTIVSSHSPGQDAEDEEA 646
Db 569 DGSANAQTLVKLPTLGPKAENSVEEKNSLNLNLPANSLKVINSVGNPPVKEKAVLIDQT 628
QY 647 TPHELQPMT-----LNGANSADSDPDKENGGLAPDGASCCQOPALHSENPEKANGLPK 702
Db 629 MGHLEESAISLSKLTSETRELTTPK-----NTREPKTTLKVEFK 671
QY 703 LMPAPLLSPEDKILETFRLSNKLKG 728
Db 672 FFKL-ALGLRKKKVVORRERLSTTVAG 696

RESULT 12
Q9FPS2
ID Q9FPS2 PRELIMINARY; PRT; 661 AA.
AC Q9FPS2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
DE Ubiquitin-specific protease 25.
GN UBP25.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20567829; PubMed=11115897;
RA Van N., Doelling J.H., Falbel T.G., Durski A.M., Vierstra R.D.;
RT "The Ubiquitin-Specific Protease Family from Arabidopsis. AtUBP1 and 2
Are Required for the Resistance to the Amino Acid Analog Canavanine.";
RL Plant Physiol. 124:1828-1843(2000).
DR EMBL; AF302673; AAG42763.1; -.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004221; P:ubiquitin thiolesterase activity; IEA.
DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
DR InterPro; IPR001394; Peptidase_C19.
DR Pfam; PF00443; UCH_1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS00235; UCH_2_3; 1.
DR Protease.
KW Protease.
SQ SEQUENCE. 661 AA; 73293 MW; D47EAD85B8D9B8E CRC64;

Query Match
Best Local Similarity 18.1%; Score 737; DB 10; Length 661;
Matches 200; Conservative 113; Mismatches 230; Indels 128; Gaps 21;

```


Db 72 SVLENLSKVIKPGNP-----GAINGFSGKNN-----TGKLVGANGHNDNGARKQAE 120
 Qy 74 -PSPQDQ-----ALGDGIAPQKVLFPBEKICLKWOQTHR-----VGAGLQNLGNT 119
 Db 121 HPNQSHHNNHHQHPTSPNELPKPKRVLYPRENIRIGWKQSERKWOVGTGMVNGT 180
 Qy 120 CFANAALQCLTYTTPPLANMLSHS-HSKTCH-AR--GFCMCTMWOAHITQALSPGVDIK 175
 Db 181 CYLNTSTIQAHLHIFALANLWVSEQAHLADCNVAPGGGCIICAMTKILLATQSNQ-SAVR 239
 Qy 176 PMFVINEMRRIARHLRFGNQDAHEFLQYTVDMQKACL---NGSNKLDHRHTQATTLVCQ 232
 Db 240 PFLYLSKLQICKHVVVGRQEDAHEFLRFLVEAMERAYLMFRVNYKELDQLVKETPLGQ 299
 Qy 233 IFGYLSRVKCLNCKGVSUFTDFPYLDITLILKAAQSVNKALEOFVPEQLDGENSKCS 292
 Db 300 IFGYLSRVKCLNCKGVSUFTDFPYLDITLILKAAQSVNKALEOFVPEQLDGENSKCS 358
 Qy 293 KCKKMWVPASKRFTTHRSNVLTSLKRFANFTGKIAKDVKYPEYLDIRPYMSQ---PNG 349
 Db 359 GCKKKVATKQFSLERAPITLCIQKRF-SMIGNKLTQKISFKSRIDLKVAARSQAQA 417
 Qy 350 EPIVYLVYAVLVHTGFGNCHAGHYFCYIKASNGLYQMDNSTVSTDIRSVLSQOAYLVFY 409
 Db 418 QPLTYRLVSMVTHLGASCHGCHYTAIGSTDGTSFYNEFDDSVYRPIAMHSVCNTNAYIMFF 477
 Qy 410 -----IRSHDVK-NGGBLTHPTSPGQSSRPVLSQRVVTKQAAPGIFGQL 456
 Db 478 ELDLQAASPAANRPNGLTNGHSTTPVPAATVSSPSPT-----RFIGFQL 524
 Qy 457 PSHMIKNPHLN-----GTGPKLDT-----PSSSMGSP 484
 Db 525 PAGGANGYTNQAOKTALQKQNOQSPONGLOLGTGKFQDTAKPPLVGAHAKGATSA 584
 Qy 485 NGSSVNRASPVNASASVQNVNRSVPIPHPKKQKITISIHNLKPVRCQSQPNLHNS 544
 Db 585 TANG--NKSSSPSSSNHKSINQOQVLPISDDIEDDMKPRTPTAQLPSMPNM--- 639
 Qy 545 SLENTFPVPSSTITNSAVQSTSNASTSVSSKTKPT-----PRSESCSQPMNCKS 597
 Db 640 -TENHTFPAKSPV-----KIQVKTPTKPLSLVPIYESASEEAEPLNPRK 686
 Qy 598 KLNSSVLVPGAESSEDSDESKGLKENGITGVSSHSPQ-----DAED 643
 Db 687 R-----PSGDSSE-SDQES---GQNGHSHKNGSHNGSHNGSHNGSHNGSHNGSH 735
 Qy 644 BEATPHELQEPMTLNGANSADSDPKPE-----NGLAPDGACQCOQPALHSENPPAK 695
 Db 736 E-----IFKSLKKSADSDDEDEEPSIQLTNGWHHPKQOS-QSQ-----SK 775
 Qy 696 ANGLPGKLMAPLLSLPEDKILETFR-----LSNK----- 725
 Db 776 APPSP-KTPPSPAVIKSTGKVTWTRNDEVDIAEDDDVVVVVVEGSPVKIPTPNKRNHPF 834
 Qy 726 --LKGSTDEMSPAER 740
 Db 835 SSSKPSIDSPATPGAKR 851

RESULT 15

Q8VZF5
 ID Q8VZF5 PRELIMINARY; PRT; 661 AA.
 AC Q8VZF5;
 DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE AT3g14400/MLN21.18.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsiis.
 OX NCBI_TaxID=3702;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Banh J., Bowser L.,
 RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
 RT "Arabidopsis cDNA clones";
 RL submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY064992; AAL5764; 1; -;
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
 DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
 DR InterPro; IPR001394; Peptidase_C19.
 DR Pfam; PF00443; UCH; 1.
 DR PROSITE; PS00972; UCH_2_1; 1.
 DR PROSITE; PS00973; UCH_2_2; 1.
 DR PROSITE; PS50235; UCH_2_3; 1.
 SQ SEQUENCE 661 AA; 73254 MW; 607DEADDBDC95647 CRC64;

Query Match 18.0%; Score 734; DB 10; Length 661;
 Best Local Similarity 29.7%; Pred. No. 1.1e-43;
 Matches 199; Conservative 114; Mismatches 230; Indels 128; Gaps 21;

Qy 98 KICLKW-----QQTHRVG--AGLQNLGNTCFANAALQCLTYTTPPLANMLSHHSKTC-- 148
 Db 4 KQSWMPVLLSKQKRNGLNLTCTVNSVLQCLFTFTPLANFCLTHKSHSCHDT 63
 Qy 149 HAEG-----FCMCTMWOAHITQALSNPGDVTKPMFVINEMRRIARHLRFGNQDAHEFLQY 204
 Db 64 YVDCERKRDPCFVIEKRIARSLSVLTTDAPNKISSCLKIFAEHFKLGRDEAEFLRY 123
 Qy 205 TVDA-----MOKACLNGSNKLDHRHTQATTLVCQIFGYLSRVKCLNCKGVSUFTDPY 257
 Db 124 VIDACHNTSLRLKRLRYNGNEPFN---GNSVVKIEFGALQSOVKCLSCGAENKADEI 179
 Qy 258 LDITLEIKAAQSVNKALEOFVPEQLDGENSKCKKMWVPASKRFTTHRSNVLTSL 317
 Db 180 MDISLEILQSSSVKESLQKPFQSBILDGNNKYRCESCEKLVATARKQMSILOAPNLIQL 239
 Qy 318 KRPFANFTGGKIADKVPYLDIRPYMSQPNRGP-IVVVLVAVLVHTGFGNCHAGHYFCYI 376
 Db 240 KRFGGIFGGKIDKA-SFGEILVLSNFMNSKASKDPQPKYKLFGLIIVHSGFSPESGHYAYV 299
 Qy 377 KASNGLYQMDNSTVSTDIRSVLSQOAYLVFYIRSHD-----VKNCGELTHPTHP 428
 Db 300 KDSLGRWYCCNDSPVLSUJTOEVLSEKAYILFFRSNQRPSAKTILVTSNGTTSHVNGC 359
 Qy 429 QSSSPRPVISOQVWVTKQAAPGFIGPQLPSHMIKNPHNLNGTGLKOTPSSSMSPNGNS 488
 Db 360 ETSNPQK-----FIGP-----LNG---FNMKPAEQSFQKGNL 389
 Qy 489 SVNRASPVNASASVQNVNRSVPIPEHPKQKITISIHNLKPVRCQSQPNLHNSLEN 548
 Db 390 ASSK-----PHK-----FIRPKPRAEQAPLENDLLSSKVEK 420
 Qy 549 -PTKPVPSSTITNSAVQSTSNASTSVSSKVTKPIPRSESCSQP--VMNGSKLNSSVLV 605
 Db 421 APLRPHAKVSI-----SVNLGAKRVSPVNGRLSFHDENIAPKANKENSVSVL 468
 Qy 606 PYGAESSEDSDESKGLGKEN-GIGTIVSSHSPQDADEBATPHELQEPMTLNGANSAD 664
 Db 469 PTKVNSGTE-----RKFGTENGGVKENGSGAPSSNNHKVLAHPHE-----RSNGSS 515
 Qy 665 SDSDPKENG LAPDGACQCOQPALHSENPPAKVANGLPKGLMAPLLSLPEDKILETFRLSN 724
 Db 516 NCGDHHKDNLHPCGNSQNGTAH---PETERGVSTTQSKGLCSSTKEDPCI-----LLR 568
 Qy 725 KLKGSTDEMSPA 735
 Db 569 KDSSRNLEEA 579

Search completed: August 10, 2004, 19:59:49
Job time : 46 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:57:24 ; Search time 21 Seconds
(without alignments)
1905.244 Million cell updates/sec

Title: US-10-049-745-4

Perfect score: 4070

Sequence: 1 MTIVDKASESSDPSAYQNP.....AAESLEPDAASLFFPSEG 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1423.5	35.0	1123	4	US-09-408-865-1
2	1127	27.7	494	3	US-09-019-095A-24
3	889.5	21.9	545	3	US-09-019-095A-38
4	846.5	20.8	521	3	US-09-019-095A-22
5	846.5	20.8	526	3	US-09-019-095A-2
6	841.5	20.7	509	3	US-09-019-095A-9
7	830.5	20.4	508	3	US-09-019-095A-8
8	678	16.7	418	3	US-09-019-095A-10
9	335	8.2	789	1	US-08-938-534-20
10	335	8.2	789	2	US-08-938-534-20
11	335	8.2	789	4	US-09-345-294-20
12	304	7.5	292	4	US-09-406-045-5
13	304	7.5	926	1	US-08-159-340A-2
14	270.5	6.6	2547	3	US-09-058-489-35
15	269.5	6.6	2555	3	US-09-058-489-36
16	247.5	6.1	242	3	US-09-019-095A-23
17	217	5.3	912	1	US-07-789-915A-8
18	217	5.3	912	1	US-08-005-002C-8
19	217	5.3	912	1	US-08-487-203A-8
20	211	5.2	291	4	US-09-167-206-22
21	193.5	4.8	372	4	US-09-406-045-1
22	191	4.7	56	3	US-09-019-095A-15
23	191	4.7	56	3	US-09-019-095A-33
24	186	4.6	462	4	US-09-166-350-18
25	184	4.5	56	3	US-09-019-095A-16
26	172	4.2	907	3	US-08-783-774-2
27	172	4.2	907	4	US-09-328-599A-1

28	172	4.2	907	5	PCT-US95-04611A-19	Sequence 19, Appl
29	170.5	4.2	1291	3	US-09-150-460B-10	Sequence 10, Appl
30	170.5	4.2	1291	3	US-09-220-641-5	Sequence 5, Appl
31	170	4.2	878	4	US-09-556-706B-2	Sequence 2, Appl
32	165.5	4.1	1581	3	US-09-110-517-2	Sequence 2, Appl
33	164	4.0	1939	4	US-09-854-856-48	Sequence 14, Appl
34	164	4.0	1999	4	US-09-854-856-16	Sequence 48, Appl
35	164	4.0	2076	4	US-09-854-856-46	Sequence 16, Appl
36	164	4.0	2136	4	US-09-854-856-14	Sequence 46, Appl
37	162	4.0	1911	4	US-09-854-856-64	Sequence 14, Appl
38	162	4.0	1971	4	US-09-854-856-32	Sequence 64, Appl
39	162	4.0	2048	4	US-09-854-856-62	Sequence 32, Appl
40	162	4.0	2108	4	US-09-854-856-30	Sequence 62, Appl
41	152.5	3.7	1290	3	US-09-150-460B-6	Sequence 30, Appl
42	151.5	3.7	875	3	US-09-150-460B-7	Sequence 7, Appl
43	149.5	3.7	2032	4	US-09-854-856-42	Sequence 42, Appl
44	149.5	3.7	2092	4	US-09-854-856-10	Sequence 10, Appl
45	149.5	3.7	2169	4	US-09-854-856-40	Sequence 40, Appl

ALIGNMENTS

RESULT 1
US-09-408-865-1
; Sequence 1, Application US/09408865A
; Patent No. 6329171
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 23484, A NOVEL HUMAN UBIQUITIN
; FILE REFERENCE: 5800-51
; CURRENT APPLICATION NUMBER: US/09/408,865A
; CURRENT FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1123
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-408-865-1

Query Match 35.0%; Score 1423.5; DB 4; Length 1123;
Best Local Similarity 37.3%; Pred. No. 1.6e-116;
Matches 347; Conservative 112; Mismatches 238; Indels 233; Gaps 25;

QY	1	MTIVDKASESSDPSAYQNPQSGSEAVSPGMDAGSAGVSSINDV-----	47
Db	1	MPIVDKLKEAL-----KPGKDSADDDGELGLASSAKKVLLOKIEFPASKFSYQL	53
QY	48	-----SNHTLSLGPVPGAVVY-SSSVDPKSPQKQDQALGDGIAPPQKVLFPSEKICLK	102
Db	54	EALKSKYVLLNPKTEGASRHKSGDDPPARRQSGSEHTVSCGDGVPAPQKVLFPTELSLR	113
QY	103	WOOTHVAGLQNLGNTCPANAALQCLTTPPLANTMLSHSEKTHAEGFCMMCTMOAH	162
Db	114	WERVFRVAGLHNLGNTCFNATIQCLTTPPLANTMLSHSEKTHAEGFCMMCTMOAH	173
QY	163	ITQALSNPGDVIKPMFVINEMRIRARHFRGQEDAEHFLQYTDAMOKACLNGSKLDR	222
Db	174	IYQAFANSNAIKPVSFIRDLKRIARHFRGQEDAEHFLRYTIDAMQKACLNGSKLDR	233
QY	223	HTQATTLVQCIFFGYLRSRVKCLNCKGVSDTDPYLDITLTKAAQSVNKALEQFVKPEQ	282
Db	234	QTQATTLVHQIFGGYLSRVKSVCKSVSDTYDPLDVALEIRQAANIYVALELFVKADV	293
QY	283	LDGENSYKSKCKKMPASKRTIHRSSNVLTSLSKRFANFTGGKIAKDVKYPEYLDIRP	342
Db	294	LSGENAYMCAKCKKVPASKRTIHRSSNVLTSLSKRFANFTGGKIAKDVKYPEYLDIRP	353
QY	343	YMSQNGEPVIVVLVAVLVHTGNCCHAGYFCYIKASGLWYOMNDSIVSTSDIRSVLSQ	402
Db	354	YMSQNGDPVITGLYAVLVHSGYSGHAGHYCYVKAASNGQWYOMNDSLVHSSNVKVLNQ	413

QY 366 NCHAGHYFCYIKASNGLWYQNDSTVSTDIRSVLSQAYVLFYIR-----SHDVKNK 418
Db 302 TSHSGHYFCVCKAGHGKWKYKMDTKVTRCDVTSVLNENAVLVFVQQANLKQVSDMPEG 361
QY 419 --GELTHPTH-----SPQSSPRPVISO-----RVVTNKQ 446
Db 362 RINEVLDPEYQLKSRKHKHKKSPFTEDLGEPCENRDKRAIKETSLGKGKVLQEVNKK 421
QY 447 AAPGF-----IGPOLPSHMIKNPPHLNLTGPKLDTSPSS--MSSPNGNSVNRASP----- 495
Db 422 AGQKHGNTKLMPOKQNHQ-KAGQNLNTEVELDLPADAIIVHQPSTANWGRDSPKENQ 480
QY 496 -----VNASASVQVNSVNR 509
Db 481 PLHNADRLTSGQPVNTWQLCR 502

RESULT 6
US-09-019-095A-9
; Sequence 9, Application US/09019095A
; Patent No. 6287858
; GENERAL INFORMATION:
; APPLICANT: D'Andrea, Alan D.
; APPLICANT: Zhu, Yuan
; TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
; TITLE OF INVENTION: Cell Growth
; FILE REFERENCE: DFCI-435P2A2
; CURRENT APPLICATION NUMBER: US/09/019,095A
; CURRENT FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US96/12884
; PRIOR FILING DATE: 1996-08-07
; PRIOR APPLICATION NUMBER: US 60/002,066
; PRIOR FILING DATE: 1995-08-09
; PRIOR APPLICATION NUMBER: US 60/019,787
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Murine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(509)
; NAME/KEY: VARIANT
; LOCATION: (1)...(509)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-019-095A-9

Query Match 20.7%; Score 841.5; DB 3; Length 509;
Best Local Similarity 49.4%; Pred. No. 1.3e-65;
Matches 162; Conservative 62; Mismatches 95; Indels 9; Gaps 4;

QY 91 KVLFPSEKICLKWOOTHRVAGLQNLGNTCFANALQCLTYTPPLANYMLSHESKTHA 150
Db 20 EVLTNGKFSLSWSPXPGCGGLQNTGNSCYLNAALQCLTHTPPLADYMLSQHSQTCCS 79
QY 151 EGFQWCMCTMOAHITQA--LSNPGDVIKPMFVINEWRRIARHLRFGNQDAHEFLQYVDA 208
Db 80 PEGCKMCAMECVTQSLXLSLXGDMVKPSQILTS-----APFKH--QDEDAHEFLMFTLET 133
QY 209 MQKACLSNGSKLDRHTQATTLVLCQIFGGYLSRVKLCNCKGVSTDFDYLDTITLIEKAAQ 268
Db 134 MHESCLQVHRSDTFPQDTSIHIHIFGGWWSQIKCLXA-GTSHTFDFDVLVDLSSAQ 192
QY 269 SVNKALEQFVKEQDGENSKYKCKKQWPKASKRFTIHRSSNVLTLKLPANFTGGKI 328
Db 193 SVNQALWDTGKSEELGENAYYCGRCROKWPASKTLHVHAPKVLILLVKRFSFTGNKL 252
QY 329 AKDVYPEYLDIRPYMSQNGEPIVLYAVLVHTGCHAGHYFCYIKASNGLWYQND 388
Db 253 DRKVSYPEFLDLKPYLSEPTGGPLPYALYAVLVHDGATSNHSGHYFCVCKAGHGKWKYKMD 312

QY 389 SIYSTDIRSVLSQAYVLFYIRSHDVK 416
Db 313 TKVTRCDVTSVLNENAVLVFVQQADLK 340

RESULT 7
US-09-019-095A-8
; Sequence 8, Application US/09019095A
; Patent No. 6287858
; GENERAL INFORMATION:
; APPLICANT: D'Andrea, Alan D.
; APPLICANT: Zhu, Yuan
; TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
; TITLE OF INVENTION: Cell Growth
; FILE REFERENCE: DFCI-435P2A2
; CURRENT APPLICATION NUMBER: US/09/019,095A
; CURRENT FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US96/12884
; PRIOR FILING DATE: 1996-08-07
; PRIOR APPLICATION NUMBER: US 60/002,066
; PRIOR FILING DATE: 1995-08-09
; PRIOR APPLICATION NUMBER: US 60/019,787
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Murine
US-09-019-095A-8

Query Match 20.4%; Score 830.5; DB 3; Length 508;
Best Local Similarity 37.0%; Pred. No. 1.2e-64;
Matches 187; Conservative 89; Mismatches 150; Indels 79; Gaps 14;

QY 67 SVPDKSPQKQDQALGDIAPPQKV--LFPSEKICLKWOOTHRVAGLQNLGNTCFANA 124
Db 4 SSPD--APFLHQDEA-----QVBEELTVNGKHSLSWSPQPGCGGLQNTGNSCYLNA 53
QY 125 ALQCLTYTPPLANYMLSHESKTHCHAGFCMCKTMOAHITQAL--SNPGDVIKPMFVINE 182
Db 54 ALQCLTHTPPLADYMLSQHSQTCCSPGCKLCAMEALVTQSLHSHSGDVNKPFSHILTS 113
QY 183 MRRIRARHLRFGNQDAHEFLQYTVDMQKACLSNGSKLDRHTQATTLVLCQIFGGYLSRV 242
Db 114 ----APFKH--QDEDAHEFLMFTLETMHESCLQVHRQSKPTSEDSSPIHDFGGWWSQI 167
QY 243 KCLNCKGVSTDFDYLDTITLIEKAAQSVNKALEQFVKEQDGENSKYKCKKQWPKASK 302
Db 168 KCLLCQGTSDTYDRFLDPLDISSAQSVKQALWDTKSEELCGDNAYYCGRCROKWPASK 227
QY 303 RFTIHRSSNVLTLKRPANFTGGKIADKVKYPEYLDIRPYMSQNGEPIVLYAVLVH 362
Db 228 TLHVHIAKPLVMVNLNRPASFTGNKLDKRVSYPEFLDLKPYLSEPTGGPLPYALYAVLVH 287
QY 363 TGFCHAGHYFCYIKASNGLWYQNDSTVSTDIRSVLSQAYVLFYIR-----SHDV 415
Db 288 DGATSHSGHYFCVCKAGHGKWKYKMDTKVTRCDVTSVLNENAVLVFVQQANLKQVSDM 347
QY 416 KNG--GELTHPTH-----SPQSSPRPVISO-----RVVT 443
Db 348 PEGRIINEVLDPEYQLKSRKHKHKKSPFTEDLGEPCENRDKRAIKETSLGKGKVLQEVN 407
QY 444 NKQAAFGF-----IGPOLPSHMIKNPPHLNLTGPKLDTSPSS--MSSPNGNSVNRASP-- 495
Db 408 HKKAGQKHGNTKLMPOKQNHQ-KAGQNLNTEVELDLPADAIIVHQPSTANWGRDSPDK 466
QY 496 -----VNASASVQVNSVNR 509
Db 467 ENQPLHNADRLTSGQPVNTWQLCR 491

RESULT 8

US-09-019-095A-10
; Sequence 10, Application US/09019095A
; Patent No. 6287858
; GENERAL INFORMATION:
; APPLICANT: D'Andrea, Alan D.
; TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
; FILE REFERENCE: DFCI-435p2A2
; CURRENT APPLICATION NUMBER: US/09/019, 095A
; PRIOR FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US96/12884
; PRIOR FILING DATE: 1996-08-07
; PRIOR APPLICATION NUMBER: US 60/002,066
; PRIOR FILING DATE: 1995-08-09
; PRIOR APPLICATION NUMBER: US 60/019,787
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 418
; TYPE: PRI
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: consensus
US-09-019-095A-10

Query Match 16.7%; Score 678; DB 3; Length 418;
Best Local Similarity 38.3%; Pred. No. 2.6e-51;
Matches 158; Conservative 73; Mismatches 114; Indels 68; Gaps 18;
QY 101 LKWOOTHRVAGLQNLGNTCFANAALQCLTYTPPLANYMLSHEHSKTCCHAEFGCMCTWQ 160
DB 26 LSW-ESPFGCGLQNTGNSCYNAALQCLTHPPPLADYMLSEHSOTCCSPGC-KCAME 83
QY 161 AHITQALSNPGDVIKPMFVINEMRRIARHLRFGNQDEAHEFLQYTVDMQKACLNGSNKL 220
DB 84 --VTQSL--GDVMPKPSILTSAPHK-----HQQDEAHEFLMFLTETHESCL-----QV 128
QY 221 DRHTQATLVCOIFGGYLSRVKCLNCKGVSDTFDPYLDITLIEKAAQSVNKALQEFVKP 280
DB 129 HRQSPDTPSIHDFGGMWRSSQIKCLGTS--TDFLDP-----LDISSAQSVQALWD--TKS 179
QY 281 EOLDGENSKCKKQKQWVPASKRFTIHRSSNVLTLSLKEFANFTGKIAKQVYKPYLDI 340
DB 180 EELG--NAYCG-CRQKMPASKTLVHIAPKVLVL---RPSAFTGNKLDKRVSYPEFLDL 233
QY 341 RPYMSQNGEPIVYVLYAVLVHTGFNCHAGHYFCYIKASNGLWYQMNDSIVSTSDIRSVL 400
DB 234 KPYLSEPTGGPLPYALYAVLVHDGAT--SSGHYFCVCKAGHGWKYMWDTKVTRCDVTSVL 292
QY 401 SQQAVLVFIR-----SHDVKNKGGLTHP-----THSPGQSSPR 434
DB 293 NENAYLVFVQOALKQVSDIMPEGREVLDPYQLKKSRRKKHKGEDRLGEVNHKAGHNTK 352
QY 435 PVISORVVNKAAPGF---IGPOLPSHMIKNPPHNLGTGPKLTPSSMS 483
DB 353 --LPQO---NHQGNLNRTEVELDLPDIVIHOPRSTANWNGSKENQPNRLTS 400

RESULT 9
US-08-431-080-20
; Sequence 20, Application US/08431080
; Patent No. 5698686
; GENERAL INFORMATION:
; APPLICANT: Gottschling, Daniel E.
; TITLE OF INVENTION: Telomerase Compositions and Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston

STATE: TEXAS
COUNTRY: UNITED STATES OF AMERICA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/431,080
FILING DATE: Concurrently Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SN 08/326,781
FILING DATE: October 20, 1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: ARCD:155/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 789 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-431-080-20

Query Match 8.2%; Score 335; DB 1; Length 789;
Best Local Similarity 23.4%; Pred. No. 1.6e-20;
Matches 112; Conservative 74; Mismatches 173; Indels 120; Gaps 16;
QY 96 SEKICLKW--QOTHRAGLQNLGNTCFANAALQCLTYTPPLANYML-----SHEHSKTC 148
DB 330 STRIVKNWGDKFTNLKPRGLNHGVTCTYNAAVQAMLHIPSIOHYLFDILMGKYDSTISK 389
QY 149 HAEGFCMCTQAHITQALSNP-----GDVIKPMFVINEMRRIARHLRFGNQDEAHEFLQ 203
DB 390 NSVSYTLAFTSKMMLPVSKNPKVSVASYINPKHLISRLDDINCMMSEWQEDSHEYFM 449
QY 204 YTVDMQKACLNGSNKLDHRHTQATLVCOIFGGYLSRVKCLNCKGVSDTFDPYLDITLIE 263
DB 450 SLMSRLQEDSVKPGHKL-----IESIIVDFGGLLKQIVTCKSCGSISKTEQPFYDLSLH 504
QY 264 IKAAG----- 268
DB 505 LKGGKKLDPNSLSDSINGTSATTTSTTSNAATKPSLSSSSSVNLNNGSPFAAASDLSS 564
QY 269 -----SVNKALQFVKPE--QLDGE--NSYKCSKCKQWVPASKRFTIHRSSNVLTLSLKRF 320
DB 565 ANRRSIEKSIKDFENPELLIKVDKEQGYVCEKCHKTNAVXSSILRAPETLLVHLKRF 624
QY 321 ANFTG---GKIAKQVYKPYLDIRPYMSQNGEPIVYVLYAVLVHTGFNCHAGHYFCYIK 377
DB 625 -RFNGTSSSKMKQAVSYPMFLDLTEY-CESKELPKYQLLSVVVHGRSLSSGHYIAHCK 682
QY 378 ASNGLWYQMNDSIVSTSDIRSVLQ-QAYVLPIRSHDVKNKGELTHPTHSFGQSSPRV 436
DB 683 QPDGGSWATYDDEYINIIISERDVLKEPNAYLLYTR-----LT-----PKSVPLPL 727
QY 437 ISORVVNKAAPGFIGPOLPSHMIKNPPHNLGTGPKLTPSSM-----SSPN 485
DB 728 AKSAAT-----GNVTSKSKQEQAVNEPN---NRPLKINSKKNRKKWKIKKGGSSPN 777
RESULT 10
US-08-938-534-20
; Sequence 20, Application US/08938534
; Patent No. 5916752

```

; GENERAL INFORMATION:
; APPLICANT: Gottschling, Daniel E.
; APPLICANT: Singer, Miriam S.
; TITLE OF INVENTION: Telomerase Compositions and Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TEXAS
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,534
; FILING DATE: 26-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/431,080
; FILING DATE:
; FILING DATE: October 20, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:155/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 789 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-938-534-20

```

```

Query Match      8.2%; Score 335; DB 2; Length 789;
Best Local Similarity 23.4%; Pred. No. 1.6e-20;
Matches 112; Conservative 74; Mismatches 173; Indels 120; Gaps 16;

Qy 96 SEKICLKW--QOTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYML-----SHEHSKTC 148
Db 330 STRIVKNGDKFTNLKPRGLNHGVTCTVNAAVQAMLHPSIQHYLFIDILMGKYDSTISK 389

Qy 149 HAEGFCMCTQAHITQALSNP-----GDVTKPMFVINEMERIAHRLFRGNOEDAHEFLQ 203
Db 390 NSVSYTTLAETSCKMWLPVSKNPKNVASYNPKHLISRLDDINCMSEWOQEDSHEYFM 449

Qy 204 YTDAMOKACLGNGKLDHRHTQATTLVCQIFGGYLRSRVKCLNCKGVSDTDFPDYLDITLE 263
Db 450 SLMSRLQEDSVPKGHKL-----IESIYDI FGGLLKQIVTCCKSGSISKTEQPFYDLSLH 504

Qy 264 IKAQAQ----- 268
Db 505 LKGGKKLDPNGLSDSDSINGTSATTTSTTNAATKPSLSSSSSSVNLNNGSPFAAASDLSS 564

Qy 269 -----SVNKALEQFVKPE--OLDGE-NSYKCSCKCKMVPASKRFTIHRSSNVLTLSLKRF 320
Db 565 ANRRPSIEKSIKDFNPELIIKVDKEKQGVCEKCHKTNTNAVKHSSILRAPETLLVHLKRF 624

Qy 321 ANFTG---GKIAKDVKYPEYLDIREYMSQNGEPIVYVLYAVLHTGFNCCHAGHYFCYIK 377
Db 625 -RFNGTSSSKMKQAVSPWELDLTEY-CESEKLPVKYQLLSVVVHEGRSLSSGHYIAHCK 682

Qy 378 ASNGLWYQNDISVSTSDIRSVLQ--QAYVLFIYIRSHDVKNGGELTHPTHPGQSSPRPV 436
Db 683 QPDGSWATYDDYINIISERDVLKPNAYLLYTR-----LT-----PKSVPLFL 727

```

```

Qy 437 ISQRVVTKQAAPGFIGPOLPSHMKNPPLHNGTGPLKDTFSSSM-----SSPN 485
Db 728 AKSAMAT-----GNVTSKSKQEQAVNEPN--NRPLKINSKKNRKKKKKKGSSPN 777

RESULT 11
US-09-345-294-20
; Sequence 20, Application US/09345294
; Patent No. 6387619
; GENERAL INFORMATION:
; APPLICANT: Gottschling, Daniel E.
; APPLICANT: Singer, Miriam S.
; TITLE OF INVENTION: Telomerase Compositions and Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TEXAS
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/345,294
; FILING DATE: 30-Jun-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/431,080
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:155/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 789 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:

US-09-345-294-20

```

```

Query Match      8.2%; Score 335; DB 4; Length 789;
Best Local Similarity 23.4%; Pred. No. 1.6e-20;
Matches 112; Conservative 74; Mismatches 173; Indels 120; Gaps 16;

Qy 96 SEKICLKW--QOTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYML-----SHEHSKTC 148
Db 330 STRIVKNGDKFTNLKPRGLNHGVTCTVNAAVQAMLHPSIQHYLFIDILMGKYDSTISK 389

Qy 149 HAEGFCMCTQAHITQALSNP-----GDVTKPMFVINEMERIAHRLFRGNOEDAHEFLQ 203
Db 390 NSVSYTTLAETSCKMWLPVSKNPKNVASYNPKHLISRLDDINCMSEWOQEDSHEYFM 449

Qy 204 YTDAMOKACLGNGKLDHRHTQATTLVCQIFGGYLRSRVKCLNCKGVSDTDFPDYLDITLE 263
Db 450 SLMSRLQEDSVPKGHKL-----IESIYDI FGGLLKQIVTCCKSGSISKTEQPFYDLSLH 504

Qy 264 IKAQAQ----- 268
Db 505 LKGGKKLDPNGLSDSDSINGTSATTTSTTNAATKPSLSSSSSSVNLNNGSPFAAASDLSS 564

Qy 269 -----SVNKALEQFVKPE--OLDGE-NSYKCSCKCKMVPASKRFTIHRSSNVLTLSLKRF 320

```

Db 565 ANRRPSIEKSIDPENPELIKVDKEQGVCEKCHKTNAVKNHSSILRAPETLLVHLKKF 624
Qy 321 ANFTG--GKIADVKYPEYLDIRYMSQNGEPIVYVLYAVLVHTGPNCHAGHYFCYIK 377
Db 625 -RFNGTSSKMKQAVSYPFMDLTRY-CESKELPVKYQLLSVVVHEGRSLSGSHYHACK 682
Qy 378 ASNGIWMQNDISVSTSDIRSVLSQ-QAYVLFYIRSHDVKNKGELTHTHSPGQSSPRPV 436
Db 683 QPDGSMATYDDDEYINISERDVLKPNAYLLYTR-----LT-----PKSVPLPL 727
Qy 437 ISQRTVINKQAAPGFIGQLPSHMKPNPHLNGTGLKMDTPSSM-----SSPN 485
Db 728 AKSAMAT-----GNVTSKSQEQAVNEPN--NRPLKINSKKNRKKKKIKKGSSN 777

RESULT 12
US-09-406-045-5
; Sequence 5, Application US/09406045
; Patent No. 6451994
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 23413, A NOVEL HUMAN UBIQUITIN
; FILE REFERENCE: 5800-50
; CURRENT APPLICATION NUMBER: US/09/406,045
; CURRENT FILING DATE: 1999-09-27
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ProDom consensus sequence
US-09-406-045-5

Query Match 7.5%; Score 304; DB 4; Length 292;
Best Local Similarity 30.2%; Pred. No. 1.7e-18;
Matches 88; Conservative 52; Mismatches 103; Indels 48; Gaps 13;
Qy 103 WQOTHRVG-AGLQNLGNTCFANAAQLCLTYTPPLANYMLSHBH-SKTHAEGFCM----- 155
Db 2 WDSKRGPGVTGLKNLGNTCYNNVSLQCLYHVPLREYFLEDEYSEVMNNEPLGMKGE 61
Qy 156 MCTMQAHIT-QALNPGD-VIKPMFVINEMRIARHLRFG-NQDAHPELOYTVD----- 207
Db 62 LATAYAKLVHQMWSNSKNKSVAPTOFLTITVGKFSQFSEGYQQDSQBFELKELQDAHED 121
Qy 208 --AMOKACLNGSNKLDHRHTOATTLV-----CQIFGGYLRSRVKCLN 246
Db 122 FNSLMEKPYVERQVK-DSNEKSTALVNVYSEAEWENHKKKENDSIITDIFQGFKSIKCP 180
Qy 247 CKGVSD-TFDPYLDITLRIKA-----AQSVNKALQFVKPEQLDGENSYKSKCKMVA 300
Db 181 CBHTSETTFEPMDLSLPSPSADNHQMLQCLSEFTKKTLEGDNKWKYCPKCKKQEA 240
Qy 301 SKRFTIHRSSNVLTLSLKRFPAN-----FTGGKIADVKYV-EYLDIRPYMSQ 346
Db 241 TKKLDIWLKPPVLVILHKKRFSYDROWGRDRDKLNTTVEFPLEDLMSPYVDK 291

RESULT 13
US-08-159-340A-2
; Sequence 2, Application US/08159340A
; Patent No. 5565352
; GENERAL INFORMATION:
; APPLICANT: Hochstrasser, Mark
; APPLICANT: Papa, Feroz
; TITLE OF INVENTION: DEUBIQUITINATING ENZYME: COMPOSITIONS
; TITLE OF INVENTION: AND METHODS
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433

; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,340A
; FILING DATE: 24-NOV-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: ARCD:112/HYL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 926 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-159-340A-2

Query Match 7.5%; Score 304; DB 1; Length 926;
Best Local Similarity 26.6%; Pred. No. 1.1e-17;
Matches 102; Conservative 57; Mismatches 126; Indels 98; Gaps 18;
Qy 112 GLQNLGNTCFANAAQLCLTYTPPLANYMLSHHSHKTC-----AEGFC-----MM 156
Db 563 GLENLGNSCYMNCIIQCLIGLTHLTQIFLDSDSYAKHININSKLGSKGILAKYFARLVHMM 622
Qy 157 CTMQAHITQALSNPGDVIKPMFV-----INEMRIARHLRFGNQEDAHEFLOYTVDAMQ 210
Db 623 YKEQVDGSKKIS-----ISPIKFKLACGSVNSLFKTA-----SQDCCQEFQCFLLDGLH 671
Qy 211 KACLN--GSN--KLDHRHTOAT-----TLVCQIFGGYLRSRV 242
Db 672 ED-LNQCGSNPPLKELSOEAEAREKLSLRATASSIEWERFLTDFSVIVDLFQGYASRL 730
Qy 243 KCLNCKGVSDTFDPYLDITLRIKAQAQSVNK-----ALRQFVKPEQLDGENSYKSKCKKM 297
Db 731 KCKVCSHTSTTYQPTVLSIPIPKNSRNNTIEDCFREFTKCNLENLEVDQMLCPHCEKR 790
Qy 298 VPASKRFTIHRSSNVLTLSLKRFPANFTGKIAKDVKYPEYLDIRPYMSQ-----PNG-- 349
Db 791 QBSTKQLTITLPRNLIVHLKRFNLL-NKNNDFFVIYFPELLDLPFWANDFGVFPFGVN 849
Qy 350 -----EPVVVLYAVLVHTGFCNHAGHYFCYIKASNGL---WYQMNDSD---IVS 392
Db 850 DDELPQIRQIPEFKYELXGVACHFG-TLYGCHYTAIVK--KGLKKGWLYFDDTKYKPVKN 906
Qy 393 TSDIRSVLSQQAYVLFYIRSHDV 415
Db 907 KAD--AINSNAVLYFHYRVYGV 926

RESULT 14
US-09-058-489-35
; Sequence 35, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; TITLE OF INVENTION: the Y Chromosome
; FILE REFERENCE: WHI97-08pA

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: August 10, 2004, 20:12:31 ; Search time 44 Seconds
(without alignments)
5557.423 Million cell updates/sec

Title: US-10-049-745-4
Perfect score: 775
Sequence: 1 MTIVDKASESDPSAYQNQP.....AAESLEEPDAAASLPFFSEGG 775

Scoring table: OLIGO Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 10

Total number of hits satisfying chosen parameters: 10

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	17	2.2	1016	13	Q802X0	Q802X0 brachydanio
2	14	1.8	548	4	Q9H9C5	Q9H9C5 homo sapien
3	12	1.5	180	11	Q8BV95	Q8BV95 mus musculu
4	12	1.5	285	4	Q8IXW9	Q8IXW9 homo sapien
5	12	1.5	829	10	Q9FJL7	Q9FJL7 arabidopsis
6	12	1.5	859	10	Q9FPS4	Q9FPS4 arabidopsis
7	11	1.4	842	10	Q84P80	Q84P80 oryza sativ
8	10	1.3	132	3	Q9HGT8	Q9HGT8 coccidioid
9	10	1.3	887	10	Q9MAQ3	Q9MAQ3 arabidopsis
10	10	1.3	901	10	Q9C585	Q9C585 arabidopsis

ALIGNMENTS

RESULT 1

```
Q802X0
ID Q802X0 PRELIMINARY; PRT; 1016 AA.
AC Q802X0;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to KIAA1453 protein (Fragment).
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB;
RA Strausberg R.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC047168; AAH47168.1; -.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0004379; F:glycylpeptide N-tetradecanoyltransferase ac. .; IEA.
DR GO; GO:0004221; P:ubiquitin thiolesterase activity; IEA.
DR GO; GO:0006499; P:N-terminal protein myristoylation; IEA.
DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR000903; Nmt.
DR InterPro; IPR001394; Peptidase_C19.
DR Pfam; PF00443; UCH; 1.
DR PROSITE; PS00290; IG_MHC; 1.
DR PROSITE; PS00976; NMT_2; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS00235; UCH_2_3; 1.
DR NON_PIR 1016 1016
FT SQ SEQUENCE 1016 AA; 113138 MW; 05AFD86206CC9DAA CRC64;

Query Match 2.2%; Score 17; DB 13; Length 1016;
Best Local Similarity 100.0%; Pred. No. 6.8e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 232 QIFGGYLSRVKCLNCK 248
Db 203 QIFGGYLSRVKCLNCK 219
|||||
|||||

RESULT 2
Q9H9C5
ID Q9H9C5 PRELIMINARY; PRT; 548 AA.
AC Q9H9C5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ12851.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshimo A.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK022913; BAB14306.1; -.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0004221; P:ubiquitin thiolesterase activity; IEA.
DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR001394; Peptidase_C19.
DR Pfam; PF00443; UCH; 1.
DR PROSITE; PS00290; IG_MHC; 1.
```

DR PROSITE; PS00972; UCH_2_1; 1.
 DR PROSITE; PS00973; UCH_2_2; 1.
 DR PROSITE; PS0235; UCH_2_3; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 548 AA; 60332 MW; 2C5ED382105AC2C6 CRC64;
 Query Match 1.8%; Score 14; DB 4; Length 548;
 Best Local Similarity 100.0%; Pred. No. 4.8e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 310 SNVLTLSLKRFPANF 323
 Db 321 SNVLTLSLKRFPANF 334
 RESULT 3
 Q8BV95
 ID Q8BV95 PRELIMINARY; PRT; 180 AA.
 AC Q8BV95;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CDNA FLJ12851 FIS.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RT "Analysis of the mouse transcriptome based on functional annotation of
 the RIKEN Genome Exploration Research Group Phase I & II Team;
 Nature 420:563-573 (2002)."
 RL Nature 420:563-573 (2002).
 DR EMBL; AK079342; BAC37614.1; -.
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
 DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
 DR InterPro; IPR001394; Peptidase_C19.
 DR Pfam; PF00443; UCH_1.
 DR PROSITE; PS00972; UCH_2_1; 1.
 SQ SEQUENCE 180 AA; 19513 MW; BC65A8C66379DEEA CRC64;
 Query Match 1.5%; Score 12; DB 11; Length 180;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 127 QCLTYTPPLANY 138
 Db 138 QCLTYTPPLANY 149
 RESULT 4
 Q8IXW9
 ID Q8IXW9 PRELIMINARY; PRT; 285 AA.
 AC Q8IXW9;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Similar to KIAA1453 protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Blood;
 RA Strausberg R.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC038983; AAH38983.1; -.
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.

DR GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
 DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
 DR InterPro; IPR001394; Peptidase_C19.
 DR Pfam; PF00443; UCH_1.
 DR PROSITE; PS00972; UCH_2_1; 1.
 DR PROSITE; PS0235; UCH_2_3; 1.
 SQ SEQUENCE 285 AA; 31791 MW; 23AA44DD2217342E CRC64;
 Query Match 1.5%; Score 12; DB 4; Length 285;
 Best Local Similarity 100.0%; Pred. No. 0.0031;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 127 QCLTYTPPLANY 138
 Db 138 QCLTYTPPLANY 149
 RESULT 5
 Q9FUL7
 ID Q9FUL7 PRELIMINARY; PRT; 829 AA.
 AC Q9FUL7;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Similarity to deubiquitinating enzyme.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=98403884; PubMed=9734815;
 RA Kotani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.
 Sequence features of the regions of 1,367,185 bp covered by 19
 Physically assigned P1 and TAC clones.";
 RL DNA Res. 5:203-216 (1998).
 DR EMBL; AB013396; BAB08869.1; -.
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
 DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
 DR InterPro; IPR001394; Peptidase_C19.
 DR Pfam; PF00443; UCH_1.
 DR PROSITE; PS00972; UCH_2_1; 1.
 DR PROSITE; PS00973; UCH_2_2; 1.
 DR PROSITE; PS0235; UCH_2_3; 1.
 SQ SEQUENCE 829 AA; 91549 MW; 71CBA02EBCAED055 CRC64;
 Query Match 1.5%; Score 12; DB 10; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0.0082;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 110 GAGLQNLGNTCF 121
 Db 106 GAGLQNLGNTCF 117
 RESULT 6
 Q9FPS4
 ID Q9FPS4 PRELIMINARY; PRT; 859 AA.
 AC Q9FPS4;
 DT 01-MAR-2001 (TrEMBLrel. 15, Created)
 DT 01-MAR-2001 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 23, Last annotation update)
 DE Ubiquitin-specific protease 23.
 GN UBP23
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.

RESULT 8

Q9HGT8 PRELIMINARY; PRT; 132 AA.

Q9HGT8 01-MAR-2001 (TrEMBLrel. 16, Created)

Q9HGT8 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

Q9HGT8 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

Q9HGT8 Ubiquitin carboxyl-terminal hydrolase (Fragment).

Q9HGT8 Coccidioides immitis.

Q9HGT8 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

Q9HGT8 Onygenales; microsporid Onygenales; Coccidioides.

Q9HGT8 NCBI_TaxID=5501;

Q9HGT8 [1]

SEQUENCE FROM N.A.

Q9HGT8 Delgado N., Cole G.T.;

Q9HGT8 "Identification of differentially expressed genes in Coccidioides

Q9HGT8 immitis.";

Q9HGT8 Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

Q9HGT8 EMBL; AF288062; AAC00799.1; -.

Q9HGT8 GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.

Q9HGT8 GO; GO:0016787; F:hydrolase activity; IEA.

Q9HGT8 GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.

Q9HGT8 GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.

Q9HGT8 InterPro; IPR001394; Peptidase_C19.

Q9HGT8 InterPro; IPR000626; Ubiquitin.

Q9HGT8 Pfam; PF00240; ubiquitin; 1.

Q9HGT8 Pfam; PF00443; UCH; 1.

Q9HGT8 SMART; SM00213; UBO; 1.

Q9HGT8 PROSITE; PS00553; UBIQUITIN_2; 1.

Q9HGT8 PROSITE; PS00972; UCH_2_1; 1.

Q9HGT8 Hydrolase.

Q9HGT8 FT NON TER 1 132

Q9HGT8 NON TER 132 132

Q9HGT8 SEQUENCE 132 AA; 14297 MW; 151651504DEB7968 CRC64;

Query Match 1.3%; Score 10; DB 3; Length 132;

Best Local Similarity 100.0%; Pred.No. 0.18;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 111 AGLQLNGTNC 120

DB 108 AGLQLNGTNC 117

|||||

|||||

RESULT 9

Q9MAQ3 PRELIMINARY; PRT; 887 AA.

Q9MAQ3 01-OCT-2000 (TrEMBLrel. 15, Created)

Q9MAQ3 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

Q9MAQ3 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

Q9MAQ3 CDS.

Q9MAQ3 F9911.5.

Q9MAQ3 Arabidopsis thaliana (Mouse-ear cross).

Q9MAQ3 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Q9MAQ3 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

Q9MAQ3 eurosids II; Brassicales; Brassicaceae; Arabidopsi.

Q9MAQ3 NCBI_TaxID=3702;

Q9MAQ3 [1]

SEQUENCE FROM N.A.

Q9MAQ3 STRAIN=cv. Columbia;

Q9MAQ3 Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,

Q9MAQ3 Altfah H., Nguyen M., Lam B., Southwick A., Bei, Buehler E., Chin C.,

Q9MAQ3 Chlou J., Choi E., Dunn P., Gonzalez A., Hwang B., Kim C., Koo T.,

Q9MAQ3 Lee J.M., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,

Q9MAQ3 Pham P., Sakano H., Schwartz J., Shinn P., Thaveri A., Toriumi M.,

Q9MAQ3 Vaysberg M., Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;

Q9MAQ3 Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

Q9MAQ3 EMBL; AC006424; AAF31287.1; -.

Q9MAQ3 FIR; C86453; C86453.

Q9MAQ3 GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.

DR GO:0004221; Fubiquitin thiolesterase activity; IEA.
 DR GO:0006512; P:ubiquitin cycle; IEA.
 DR GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
 DR InterPro; IPR006615; DUSP.
 DR InterPro; IPR001394; Peptidase_C19.
 DR Pfam; PF00443; UCH; 1.
 DR SMART; SM00695; DUSP; 1.
 DR PROSITE; PS00972; UCH_2_1; 1.
 DR PROSITE; PS00973; UCH_2_2; 1.
 DR PROSITE; PS0235; UCH_2_3; 1.
 SQ SEQUENCE 887 AA; 100619 MW; 419F8E5E067FA7B2 CRC64;

Query Match 1.3%; Score 10; DB 10; Length 887;
 Best Local Similarity 100.0%; Pred.No.1;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 GLQNLGNTCF 121
 |||||
 Db 297 GLQNLGNTCF 306

RESULT 10

Q9C585 ID Q9C585 PRELIMINARY; PRT; 901 AA.
 AC Q9C585;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Ubiquitin-specific protease-like protein.
 GN AT5G22030.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
 RA Rudd S., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL589883; CAC34496.1; -
 DR GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO:0008233; F:peptidase activity; IEA.
 DR GO:0004221; Fubiquitin thiolesterase activity; IEA.
 DR GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
 DR InterPro; IPR001394; Peptidase_C19.
 DR Pfam; PF00443; UCH; 1.
 DR PROSITE; PS00972; UCH_2_1; 1.
 DR PROSITE; PS00973; UCH_2_2; 1.
 DR PROSITE; PS0235; UCH_2_3; 1.
 KW Protease.
 SQ SEQUENCE 901 AA; 101704 MW; 5778D09590C87886 CRC64;

Query Match 1.3%; Score 10; DB 10; Length 901;
 Best Local Similarity 100.0%; Pred.No.1;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 GLQNLGNTCF 121
 |||||
 Db 310 GLQNLGNTCF 319

Search completed: August 10, 2004, 20:16:37
 Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 16, 2004, 11:33:11 ; Search time 14896 Seconds
(without alignments)
7018.220 Million cell updates/sec

Title: US-10-049-745-31

Perfect score: 2412
Sequence: 1 gggctgtgtgcccgcggc.....agggttaatttagcttgca 2412

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg_hum.*

31: em.htg_inv.*

32: em.htg_other.*

33: em.htg_mus.*

34: em.htg_pln.*

35: em.htg_rod.*

36: em.htg_mam.*

37: em.htg_vrt.*

38: em_sy.*

39: em.htgo_hum.*

40: em.htgo_mus.*

41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2412	100.0	2412	6	AX083128	Sequence
2	2371.6	98.3	4659	6	AX786887	Sequence
3	2365	98.1	3660	6	AX882263	Sequence
4	2365	98.1	3660	6	BD159625	Primer fo
5	2365	98.1	3660	6	AK022759	Homo sapi
6	2346.6	97.3	3669	9	BC060846	Homo sapi
7	2339.2	97.0	2347	6	AX107851	Sequence
8	1675.6	69.5	1679	6	AX364931	Sequence
C 9	727.2	30.1	152927	9	AC004895	Homo sapi
10	727.2	30.1	154791	2	AC069393	Homo sapi
C 11	714.4	29.6	217022	2	AC146146	Pan trogl
12	613.4	25.4	3075	5	BC057482	Danio rer
13	613.4	25.4	3274	5	BC047168	Danio rer
14	474.6	19.7	722	6	AR264045	Sequence
15	468	19.4	521	6	AX867589	Sequence
16	468	19.4	521	6	BD147651	Primer fo
17	467	19.4	2331	5	BC057470	Danio rer
18	450.8	18.7	3006	10	AK129363	Mus muscu
C 19	447.8	18.6	157080	2	AC083981	Homo sapi
20	447.8	18.6	166385	2	AC074340	Homo sapi
C 21	447.8	18.6	178401	9	AC134395	Homo sapi
C 22	446.2	18.5	172084	9	AC134683	Homo sapi
23	444.6	18.4	178000	9	AC130366	Homo sapi
C 24	443	18.4	35671	2	AC143749	Macaca mu
25	443	18.4	195488	9	AC130379	Homo sapi
C 26	441.4	18.3	158209	9	AC105214	Homo sapi
C 27	437.6	18.1	183855	2	AC022770	Homo sapi
28	436.2	18.1	183855	2	AC022770	Homo sapi
29	436	18.1	172366	9	AC108519	Homo sapi
C 30	435	18.0	167440	2	AC130346	Homo sapi
31	435	18.0	171474	9	AC130360	Homo sapi
32	435	18.0	171639	9	AC130365	Homo sapi
33	435	18.0	173255	2	AC130338	Homo sapi
34	435	18.0	187095	2	AC130338	Homo sapi
35	435	18.0	189043	2	AC130349	Homo sapi
36	435	18.0	190031	9	AC131269	Homo sapi
37	435	18.0	295854	9	AF228730	Homo sapi
C 38	434.6	18.0	43174	9	AC144862	Homo sapi
C 39	434.6	18.0	246224	9	AC116655	Homo sapi
C 40	433.4	18.0	188884	9	AC130367	Homo sapi
C 41	433	18.0	4746	6	E43116	Human deubi
42	433	18.0	4752	9	HUMANAA	Human DNA
43	431.2	17.9	130529	9	AC117179	Homo sapi
C 44	430.4	17.8	176050	2	AC131279	Homo sapi
45	430.2	17.8	1683	6	BD235215	Mammalian

ALIGNMENTS

RESULT 1
AX083128
LOCUS AX083128 2412 bp DNA linear PAT 28-FEB-2001
DEFINITION Sequence 31 from Patent WO0110903.
ACCESSION AX083128
VERSION AX083128.1 GI:13185035
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Yue,H., Lal,P., Tang,Y.T., Bandman,O., Baughn,M.R., Azimzai,Y.,
Lu,D.A. and Yang,J.
TITLE Proteases and protease inhibitors

Pred. No. is the number of results predicted by chance to have a


```
1201 AAAGCTAGCAATGGCTCTGGTATCAATGAATGACTCCATTGTATCTACAGTGATTT 1260
1283 AAAGCTAGCAATGGCTCTGGTATCAATGAATGACTCCATTGTATCTACAGTGATTT 1342
1261 AGATCGGTACTCAGCAACAAGCTTATGTCTCTTTTATATCAGGTCCCATGATGAAA 1320
1343 AGATCGGTACTCAGCAACAAGCTTATGTCTCTTTTATATCAGGTCCCATGATGAAA 1402
1321 AATGAGGTGAATCTTACTCATCCACCATAGCCCGCCAGTCTCTCCCGCCCGCTC 1380
1403 AATGAGGTGAATCTTACTCATCCACCATAGCCCGCCAGTCTCTCCCGCCCGCTC 1462
1381 ATCAGTCAGCGGTGTGACCAACAAACAGGCTCGCCAGGCTTTATCGGACCAAGCTT 1440
1463 ATCAGTCAGCGGTGTGACCAACAAACAGGCTCGCCAGGCTTTATCGGACCAAGCTT 1522
1441 CCCTCTCATGATAAAGAACTCACTCACTTAAATGGAGCTGGACCAATGGAAGACAG 1500
1523 CCCTCTCATGATAAAGAACTCACTCACTTAAATGGAGCTGGACCAATGGAAGACAG 1582
1501 CCAAGCAGTTCATGTCGAGTCTTAACGGGAATCAAGTGTCAACAGGCTAGTCTGTT 1560
1583 CCAAGCAGTTCATGTCGAGTCTTAACGGGAATCAAGTGTCAACAGGCTAGTCTGTT 1642
1561 AATGCTTCACTTCTGTCACAAACTGTCAGTAAATAGGTCCTCAGTATCCAGAACT 1620
1643 AATGCTTCACTTCTGTCACAAACTGTCAGTAAATAGGTCCTCAGTATCCAGAACT 1702
1621 CCTAAGAAACAAAAATTAACATCAGTATTCACAACTGTCCTGTCGCCAGCTGTGAG 1680
1703 CCTAAGAAACAAAAATTAACATCAGTATTCACAACTGTCCTGTCGCCAGCTGTGAG 1762
1681 TCTCAACCTAACCTTCATAGTAATCTTTGGAGAACCTTCCAGCCCGCTTCCCTCTCT 1740
1763 TCTCAACCTAACCTTCATAGTAATCTTTGGAGAACCTTCCAGCCCGCTTCCCTCTCT 1822
1741 ACATTTACAAATTCAGTACAGTCTACTCGAAGCATCTACAGTGTGAGTTTCTAGT 1800
1823 ACCATTACCAATTCAGTACAGTCTACTCGAAGCATCTACAGTGTGAGTTTCTAGT 1882
1801 AAAGTAACAAAAACCGATCCCGCCAGTGAATCTGCTCCAGCCCGCTGATGAATGCAAA 1860
1883 AAAGTAACAAAAACCGATCCCGCCAGTGAATCTGCTCCAGCCCGCTGATGAATGCAAA 1942
1861 TCCAAGCTGAATCCAGCGTGTGTTGCCCTATGCGCGCCAGTCTCTGAGGACTCTGAC 1920
1943 TCCAAGCTGAATCCAGCGTGTGTTGCCCTATGCGCGCCAGTCTCTGAGGACTCTGAC 2002
1921 GAGGAGTCAAAAGGGCTGGGCAAGGAGATGGGATTTGGTACGATTTGAGCTCCCACTCT 1980
2003 GAGGAGTCAAAAGGGCTGGGCAAGGAGATGGGATTTGGTACGATTTGAGCTCCCACTCT 2062
1981 CCCGCCCAAGATCCGAAGATGAGGAGCCACTCCGCAACAGCTTCAAGAACCCATGACC 2040
2063 CCCGCCCAAGATCCGAGAGATGAGGAGCCACTCCGCAACAGCTTCAAGAACCCATGACC 2122
2041 CTAAACGGTCTAATAGTCAGACAGCAAGTGAACCCGAAAGAAAACGGCTTAGCCCT 2100
2123 CTAAACGGTCTAATAGTCAGACAGCAAGTGAACCCGAAAGAAAACGGCTTAGCCCT 2182
2101 GATGTTCCAGCTCCGCAAGGCGAGCTGCGCTGCACTCAGAAATCCCTTTGCTTAAGGCA 2160
2183 GATGTTCCAGCTCCGCAAGGCGAGCTGCGCTGCACTCAGAAATCCCTTTGCTTAAGGCA 2242
2161 AACGGTCTCTCTGGAAGTTGATGCTGCTCTTTGCTGCTCTCCAGAAAGCAAAATC 2220
2243 AACGGTCTCTCTGGAAGTTGATGCTGCTCTTTGCTGCTCTCCAGAAAGCAAAATC 2302
2221 TTAGAGACCTTCAGGCTTAGCAACAACTGAAGGCTCGAGGATGAATGAGTGCACCT 2280
2303 TTAGAGACCTTCAGGCTTAGCAACAACTGAAGGCTCGAGGATGAATGAGTGCACCT 2362
2281 GGAGCAGAGAGGGGCCCTCCCGAGGACCGCAGCCGAGCTCAGCTGGCAGCCCGGCC 2340
```

```
2363 GGAGCAGAGAGGGCCCTCCCGAGGACCGCGAGCGCTCAGCTGGCAGCCCCGCC 2422
2341 GCCGAATCCCTGGAGGCGCAGATGCGCGCCCAAGCTT 2378
2423 GCCGAATCCCTGGAGGAGCAGATGCGCGCCCGCGCT 2460

RESULT 3
LOCUS AX882263 3660 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 17168 from Patent EP1074617.
ACCESSION AX882263
VERSION AX882263.1 GI:40037087
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
TITLE Primers for synthesising full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 17168 07-FEB-2001; (JP)
Research Association for Biotechnology
FEATURES
Location/Qualifiers
1..3660
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
68..>3658
/note="unnamed protein product"
/codon_start=1
/protein_id="CAE92364.1"
/db_xref="GI:40037088"
/translation="MTIVDKASESSDPYQNPQGSSEAVSPQMDMAGSAGVAVSSLL
NDVSNHISLGPVPGAVVYSSSSVPDKSPQKQOALGDGIAPQKVLFPSEKICLK
WQTHRVGAGLQNLGNTCFANAALQCLTTPPLANTMLSHEHSKTCCHAGFCMCTMQ
AHTOALNPGDVIKPMFVINEMRIARHREFGNQEDAEHFLQYTDVAKMOKLNSN
KLDRTCTATLVCQIFGGLRSKVKLCKGVSDDPDDYLDITLLETKAQSVMKALEQ
FVKPQLDGENSYKSCCKKMPASKRFTIHRSSNVLTLSLKRFANFTGCKIAKDVY
PYLDITFMSQNGEPIVYLVAVLHTGFNCHAGHYFCYTKASNLWYOMNDSIVS
TSDIRSVLSQOAYVLFYIRSHDVKNKGELTHPSPGQSSPRPVSQVVTNKQAAPG
FTGPQLSHMINKPPLNGTGLPKDTPSSSSPNSSSVNRASPVNAGSASVQNSVN
RSVITPHPKQKQITISIHKLVPQCSQPNLHNSNLTPTKPSSTITNSAVST
SNASTMSVSKVTPIPRSESCSQPMVNGSKLNSVLVYPGAESESDDEESKGLGK
ENGIGTIVSSHSPGDAEDEATPELQPTTLINGANSADSDPKENGLAPGASQ
GQALHSENPFAKANGLPGLKMPAPLLSLPEDKILETFLSNLKGSTDMSPAGAE
GPEDRDAPQPSGPAASELEBPDAAGLSSTKKAAPPDPDPTPATKEGWEAMAVAP
EPPPPAGEDIIVGDTAPPDLCDPSLTGDASPLSQDAKGMIAEGPRDSALAAPEGLS
PAPPARSEPECEOPLLVHPSGDHARDQDPQSGLAPEAEAREPPAPVLDMAVAGHPG
DAEPSPCERVEDAAAPKAPGSPAKEKIGSLRKVDHGRVSRSSSGSPAPRESRK
TEGHRHRRRTCPREDRODRHAPHHPGHGDRLSPERRSLGRCSSHHSRHSRGSBL
DWVRHHITGERGWRKFPYDRPWDRCRYHIDRYALVAARDKPKPHGGRHERAGL
HERPHKHNRGRGCEPARERHRRPSSPRAGAPHALAPHPDRFSDHRTALVAGDCN
LSDRPHHENGSKSRKRHSDSVENSDSHVEKKARRSEOKPLEBPKAKKHKK"
```

ORIGIN

```
Query Match 98.1%; Score 2365; DB 6; Length 3660;
Best Local Similarity 99.8%;
Matches 2368; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 6 GTGTGCGCGCGCGCGCGCGCGCGCGGAGGAGGAGGAGCGCGCGCGCGCGCGCGCGCTCAGAG 65
Db 1 GTGTGCGCGCGCGCGCGCGCGCGCGCGGAGGAGGAGGAGCGCGCGCGCGCGCGCGCTCAGAG 60

Qy 66 TTGAACAATGACCATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTATCAGAA 125
Db 61 TTGAACAATGACCATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTATCAGAA 120

Qy 126 TCACCTGGCAGCTCCCGAGGCGAGTCTCACCTGGAGACATGGATCAGGCTTCTGCCAGCTG 185
Db 121 TCACCTGGCAGCTCCCGAGGCGAGTCTCACCTGGAGACATGGATCAGGCTTCTGCCAGCTG 180
```

QY	186	GGGTGCTGTGCTTTCATTTGAATGATGTCTCAAAATCACACACTTCTTTTAGGACCAATACC	245
Db	181	GGGTGCTGTGCTTTCATTTGAATGATGTCTCAAAATCACACACTTCTTTTAGGACCAATACC	240
QY	246	TGGTGTGTAGTTTATTCGAGTTTCATCTGTACCTGATAAAATCAAAACCATCACACAAAA	305
Db	241	TGGTGTGTAGTTTATTCGAGTTTCATCTGTACCTGATAAAATCAAAACCATCACACAAAA	300
QY	306	GGATCAAGCCCTTAGTGATGTCGCTCTCCACAGAAAGTTCTTTTCCCATCTGAGAA	365
Db	301	GGATCAAGCCCTTAGTGATGTCGCTCTCCACAGAAAGTTCTTTTCCCATCTGAGAA	360
QY	366	GATTTGTCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGCTCCAGATTTTGGGCAA	425
Db	361	GATTTGTCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGCTCCAGATTTTGGGCAA	420
QY	426	TACCTGTTTTGCCAATGCAGCACTGCAGTGTTTAACTTACACACACCTCTTGGCAATTA	485
Db	421	TACCTGTTTTGCCAATGCAGCACTGCAGTGTTTAACTTACACACCTCTTGGCAATTA	480
QY	486	CATGCTATCATGAAACACTTCCAAAACATGTCATGCGAGAGCTTTTGTATGATGTATC	545
Db	481	CATGCTATCATGAAACACTTCCAAAACATGTCATGCGAGAGCTTTTGTATGATGTATC	540
QY	546	AATGCAAGCACATATTACCCAGGCACCTCAGTAATCTCTGGGGACGTTATTAACCAATGTT	605
Db	541	AATGCAAGCACATATTACCCAGGCACCTCAGTAATCTCTGGGGACGTTATTAACCAATGTT	600
QY	606	TGTCATCAATAGATGCGGGCTATAGTAGGCACCTCCGTTTTTGGAAACCAAGAAATGC	665
Db	601	TGTCATCAATAGATGCGGGCTATAGTAGGCACCTCCGTTTTTGGAAACCAAGAAATGC	660
QY	666	CCATGAATTCCTCAATACACTGTTGATGCTATGCGAAGAGCATGCTTGAATGCGACAA	725
Db	661	CCATGAATTCCTCAATACACTGTTGATGCTATGCGAAGAGCATGCTTGAATGCGACAA	720
QY	726	TAAATTAGACAGACACACCCAGGCCACCTCTGTTTGTGAGTATTTGGAGATACCT	785
Db	721	TAAATTAGACAGACACACCCAGGCCACCTCTGTTTGTGAGTATTTGGAGATACCT	780
QY	786	AAGATCTAGAGTCAAAATGTTTAAATTCGAAGGGGCTTTCAGATATTTTGATCCATATCT	845
Db	781	AAGATCTAGAGTCAAAATGTTTAAATTCGAAGGGGCTTTCAGATATTTTGATCCATATCT	840
QY	846	TGATATAACATTGGAGATAAAGCTGCTCAGAGTGTCAACAGGCATTTGGAGCATTTGT	905
Db	841	TGATATAACATTGGAGATAAAGCTGCTCAGAGTGTCAACAGGCATTTGGAGCATTTGT	900
QY	906	GAAGCCGGAACAGCTTGATGAGAAACTCGTACAAGTCGAGCAAGTGTAAAGATGTT	965
Db	901	GAAGCCGGAACAGCTTGATGAGAAACTCGTACAAGTCGAGCAAGTGTAAAGATGTT	960
QY	966	TCAGCTTCAAGAGGTTCACTATCCATAGATCCTCTAAATGTTCTTACACTTTCTCTGAA	1025
Db	961	TCAGCTTCAAGAGGTTCACTATCCATAGATCCTCTAAATGTTCTTACACTTTCTCTGAA	1020
QY	1026	ACGTTTTGCAAAATTTACCGGTGGAATAAATGCTTAAGATGTGAAPACCTCAGTATCT	1085
Db	1021	ACGTTTTGCAAAATTTACCGGTGGAATAAATGCTTAAGATGTGAAPACCTCAGTATCT	1080
QY	1086	TGATATTTGGCCATATATGTTCTCAACCCAGGAGGCCAAATGTCTACGCTTTGTATGC	1145
Db	1081	TGATATTTGGCCATATATGTTCTCAACCCAGGAGGCCAAATGTCTACGCTTTGTATGC	1140
QY	1146	AGTGTGCTCCACACTGGTTTTTAATTTGCCATGCTGGCCATTACTTCTGTACATAAAGC	1205
Db	1141	AGTGTGCTCCACACTGGTTTTTAATTTGCCATGCTGGCCATTACTTCTGTACATAAAGC	1200
QY	1206	TAGCAATGGCTCTGGTATCAAAATGAATGACTCCAATGTATCTACCAAGTATTTAGATC	1265
Db	1201	TAGCAATGGCTCTGGTATCAAAATGAATGACTCCAATGTATCTACCAAGTATTTAGATC	1260

QY	1266	GGTACTCAGCCAAACAAGCCTATGTCTCTTTTATATCAGGTCCCATGATGTGAAAAATGG	1325
Db	1261	GGTACTCAGCCAAACAAGCCTATGTCTCTTTTATATCAGGTCCCATGATGTGAAAAATGG	1320
QY	1326	AGGTGAATCTACTCATCCCAATAGCCCGGCAGTCTCTCCCGCCCGTCATCAG	1385
Db	1321	AGGTGAATCTACTCATCCCAATAGCCCGGCAGTCTCTCCCGCCCGTCATCAG	1380
QY	1386	TCAGCGGGTGTCCAAACAAACAGGCTGCGCCAGGCTTTATTCGGACCAACAGCTTCCCTC	1445
Db	1381	TCAGCGGGTGTCCAAACAAACAGGCTGCGCCAGGCTTTATTCGGACCAACAGCTTCCCTC	1440
QY	1446	TCACATGATAAAGAAATCCACCTCATTTAAATGGGACTGGAACATTTGAAAGACAGCCCAAG	1505
Db	1441	TCACATGATAAAGAAATCCACCTCATTTAAATGGGACTGGAACATTTGAAAGACAGCCCAAG	1500
QY	1506	CAGTTCATGTCGAGTCTTAACGGGAATTCAGAGTGTCAACAGGGCTAGTCTGTTTAAATGC	1565
Db	1501	CAGTTCATGTCGAGTCTTAACGGGAATTCAGAGTGTCAACAGGGCTAGTCTGTTTAAATGC	1560
QY	1566	TTACAGTCTCTGTCCAAAACCTGGTCAGTTAAATAGGTCTCTCAGTGATCCAGAACATCTCTAA	1625
Db	1561	TTACAGTCTCTGTCCAAAACCTGGTCAGTTAAATAGGTCTCTCAGTGATCCAGAACATCTCTAA	1620
QY	1626	GAAACAAAATAATTACATCAGTATTCACACAAAGTTGCTGCTTTCGCCAGTGTCTAGTCTCA	1685
Db	1621	GAAACAAAATAATTACATCAGTATTCACACAAAGTTGCTGCTTTCGCCAGTGTCTAGTCTCA	1680
QY	1686	ACCTAACCTTTCATAGTAATTTTGGAGAAACCTTACCAAGCCGCTTCCCTCTCTTCCCAT	1745
Db	1681	ACCTAACCTTTCATAGTAATTTTGGAGAAACCTTACCAAGCCGCTTCCCTCTCTTCCCAT	1740
QY	1746	TACCAATCTCGAGTACAGTCTACCTCGAAACGCTATCAGATGTCTAGTCTTCTAGTAAAGT	1805
Db	1741	TACCAATCTCGAGTACAGTCTACCTCGAAACGCTATCAGATGTCTAGTCTTCTAGTAAAGT	1800
QY	1806	AACAAAACCGATCCCGCGAGTGAATCCTGCTCCAGCCGCTGATGAATGGCAAAATCCAA	1865
Db	1801	AACAAAACCGATCCCGCGAGTGAATCCTGCTCCAGCCGCTGATGAATGGCAAAATCCAA	1860
QY	1866	GCTGAACTCCAGCGTGTGCTGCTTATGCGCGGAGTCTCTGAGGACTCTGACGAGGA	1925
Db	1861	GCTGAACTCCAGCGTGTGCTGCTTATGCGCGGAGTCTCTGAGGACTCTGACGAGGA	1920
QY	1926	GTCAAAGGGCTGGGCAAGGAGAAATGGATTTGGTACGAATTTGTGAGCTTCCACTCTCCCG	1985
Db	1921	GTCAAAGGGCTGGGCAAGGAGAAATGGATTTGGTACGAATTTGTGAGCTTCCACTCTCCCG	1980
QY	1986	CCAGATGCCGAAGATGAGAGGCCACTCCGCAAGAGCTTCAAGAACCCATGACCCCTAAA	2045
Db	1981	CCAGATGCCGAAGATGAGAGGCCACTCCGCAAGAGCTTCAAGAACCCATGACCCCTAAA	2040
QY	2046	CGGTGCTTAATAGTGACAGACGACAGTGAACCGAAAGAAACGGCTTACGCTGATGG	2105
Db	2041	CGGTGCTTAATAGTGACAGACGACAGTGAACCGAAAGAAACGGCTTACGCTGATGG	2100
QY	2106	TGCCAGTGTCAAAGGCGAGCTTCCCTTGCACTCAGAAAAATCCCTTTTCTAAGGCAACGG	2165
Db	2101	TGCCAGTGTCAAAGGCGAGCTTCCCTTGCACTCAGAAAAATCCCTTTTCTAAGGCAACGG	2160
QY	2166	TCTTCTCTGGAAGTTGATGCTGCTCTTTGCTGTCTCTCCCAAGAGAATAAATCTTTAGA	2225
Db	2161	TCTTCTCTGGAAGTTGATGCTGCTCTTTGCTGTCTCTCCCAAGAGAATAAATCTTTAGA	2220
QY	2226	GACCTTTCAGCTTGTAGCAAACTGAAAGCTCGAGGATGAAATGAGTGCACCTGAGC	2285
Db	2221	GACCTTTCAGCTTGTAGCAAACTGAAAGCTCGAGGATGAAATGAGTGCACCTGAGC	2280
QY	2286	AGAGGGGGCCCTCCCGAGGACCGCGACGCTCAGAGCTTGGCAGCCCGCCCGCGA	2345
Db	2281	AGAGGGGGCCCTCCCGAGGACCGCGACGCTCAGAGCTTGGCAGCCCGCCCGCGA	2340
QY	2346	ATCCCTGGAGGACCAAGATGGCGCCCGCAAGCTT	2378

[illegible]

REFERENCE
AUTHORS

1 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

2 Isogai, T., Ota, T., Hayaishi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H., Wagatsuma, M., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K., Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahari, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T.

TITLE
JOURNAL

NEO human cDNA sequencing project

UNPUBLISHED

2 (bases 1 to 3660)

Isogai, T. and Otsuki, T.

Direct Submission

Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

NEO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection

Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

FEATURES
source

Location/Qualifiers

1..3660

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="NT2RPI000522"

/cell_line="NT2"

/cell_type="teratocarcinoma"

/clone_lib="NT2RPI"

/notes="cloning vector: pUC19FL3-mRNA from NT2 neuronal precursor cells after 48-hours retinoic acid (RA) induction."

68..>3660

/note="unnamed protein product"

/codon_start=1

/protein_id="BAB14232.1"

/db_xref="GI:10434352"

/translation="MTIVDKASESDPSAYQNPQGSSEAVSPGMDAGSASMGAVSSLLNDVSNHTLSLGPVPGAVYSSSPDKSPQKQALGGIAPPOKVLFPEEKI CLKWQTHRVAGLQNLGNMTFCANALQCLTYTPLYANLMSHEHSKTCHEAGFCNMCTMOKIDRTALSNPGVDIPKPMFINAMRLIARFRFCNOBDAHEFLQYTVDMQKACLNGSNFKVPEQLQDGENSYKGSCKKKWPKAKRFTIHRSSNVLTLSLRPNFLTGGKTKADVKYPSYDLIRPVMSPNGEPIVVLVAVLVHVGFNCHAGHYFCYIKASNLGQYQNDSTVITSDIRSLVQQAVLVFIIRSHDVKNQGGELTHPTSPGSSPRVTSQRVVTKNQKAPGFIGQDPQSHMIKNPPLHMLTGLKDTPPSSSNPNSSNPNRSPASVQNMVNSRSGVTEHPKSKQKTIETSIHNKLVPVQCQSPNLHNSLENTPKVPSPSTITNSAVQSTSNASTVSSVSKYKTIPRASECSQPMWNGSKLNSVLVPYGAESSEDESDESKGLIGENGIQITIVSSHSPQADAEATPELOPPTLNGANSADSDSPKENGSLADPGASCGQPALHSPNPFKAGLQPKMLPALSIPDKIETFLSLNKLKSTDEMSAPGAERZEPEDRDAEPQGSFAASLEBPDAAGLSLTPKTPPTDPCTPATKGCWAMAVAPGEPPSAGEDIVGDTAPDLCDPSGLTDGASPSIQDAMGAETGPRDGSALAPRGLSPAPPASREPCQQLLVHSGDHDADQPSQICAPAEAREPPAPVLDDMAPAGPGLSDAEPGSGRVEDDAAPGAPSPAKEKIGSLRKVDRGHSRRSSSGEPARESRKTEGHRHRRRTCPREPRDQRPAPHPHCHGDRLS PGERRSLGCSHHSHRHSRGVELDWVRHHVTEGRGWGREKYPEDRPNKDRVYLAARDWKPFGHSGREHERAGLHERPHKDHNRGRCCEPARERHRSPKAGAPALAPDFRSHDRTALVAGDNCNLSRDFHNRGNSKRRHSDSVENSDSHVEKAKASQKDPLEEKPAKKHYS"

CDS

ORIGIN

Query Match	98.1%	Score 2365;	DB 9;	Length 3660;
Best Local Similarity	99.8%	Pred. No. 0;		

[illegible]

[illegible]

QY 306 GGATCAAGCCCTAGGTGATGGATGCTCTCTCCACAGAAAGTTCTTTTCCCATCTGAGAA 365
 Db 296 GGATCAAGCCCTAGGTGATGGATGCTCTCTCCACAGAAAGTTCTTTTCCCATCTGAGAA 355
 QY 366 GATTTGTCTTAAGTGGCAACAAACTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGGCAA 425
 Db 356 GATTTGTCTTAAGTGGCAACAAACTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGGCAA 415
 QY 426 TACCTGTTTGGCAATGCGAGCACTGAGTGTGTTAACTACACACACCTCTTTGCCAAATTA 485
 Db 416 TACCTGTTTGGCAATGCGAGCACTGAGTGTGTTAACTACACACACCTCTTTGCCAAATTA 475
 QY 486 CATGCTATCAATGAACACTCCAAACATGTGATGAGAGGCTTTTGTATGATGTTAC 545
 Db 476 CATGCTATCAATGAACACTCCAAACATGTGATGAGAGGCTTTTGTATGATGTTAC 535
 QY 546 AATGCAAGCACATATTTACCCAGGCACTAGTAACTCTGGGAGCGTTATTAAACCAATGTT 605
 Db 536 AATGCAAGCACATATTTACCCAGGCACTAGTAACTCTGGGAGCGTTATTAAACCAATGTT 595
 QY 606 TGTCTATCAATGAGATGCGGCGTATAGCTAGGCACTCTCGTTTGGAAACCAAGAGATGC 665
 Db 596 TGTCTATCAATGAGATGCGGCGTATAGCTAGGCACTCTCGTTTGGAAACCAAGAGATGC 655
 QY 666 CCATGAAATTCCTTCAATACACTGTTGATGCTATGAGAGGATGCTTGAATGGAGCAA 725
 Db 656 CCATGAAATTCCTTCAATACACTGTTGATGCTATGAGAGGATGCTTGAATGGAGCAA 715
 QY 726 TAAATTTAGACAGACACACCCAGGCACTCTCTGTTGTTGTCAGATATTTGGAGGATACCT 785
 Db 716 TAAATTTAGACAGACACACCCAGGCACTCTCTGTTGTTGTCAGATATTTGGAGGATACCT 775
 QY 786 AAGATCTAGAGTCAAAATGTTTAAATTTGAAGGCGTTTCAGATATCTTTGATCCATATCT 845
 Db 776 AAGATCTAGAGTCAAAATGTTTAAATTTGAAGGCGTTTCAGATATCTTTGATCCATATCT 835
 QY 846 TGATATTAACATTTGAGATTAAGGCTGCTCAGAGTGTCAAGGCACTTTGGAGCAGTTGT 905
 Db 836 TGATATTAACATTTGAGATTAAGGCTGCTCAGAGTGTCAAGGCACTTTGGAGCAGTTGT 895
 QY 906 GAAGCCGGAACAGCTTATGAGAGAAACTCGTCAAGTGCAGCAAGTGTAAAAAGATGTT 965
 Db 896 GAAGCCGGAACAGCTTATGAGAGAAACTCGTCAAGTGCAGCAAGTGTAAAAAGATGTT 955
 QY 966 TCAGAGCTTCAAGAGGTTCACTATCCATATGCTCTTAATGTTCTTACATCTTCTCGAA 1025
 Db 956 TCAGAGCTTCAAGAGGTTCACTATCCATATGCTCTTAATGTTCTTACATCTTCTCGAA 1015
 QY 1026 ACGTTTGGCAATTTTACCGGTGGAATAATGCTTAAGGATGTGAATACCCCTGAGTATCT 1085
 Db 1016 ACGTTTGGCAATTTTACCGGTGGAATAATGCTTAAGGATGTGAATACCCCTGAGTATCT 1075
 QY 1086 TGATATTTGGCCATATATGCTCAACCCAGGAGAGCCAAATGCTACGTTCTTGTATGC 1145
 Db 1076 TGATATTTGGCCATATATGCTCAACCCAGGAGAGCCAAATGCTACGTTCTTGTATGC 1135
 QY 1146 AGTCTGGTCCACACTGGTTTAAATGCGCATGCTGGCCATTAATCTGCTACATAAAGC 1205
 Db 1136 AGTCTGGTCCACACTGGTTTAAATGCGCATGCTGGCCATTAATCTGCTACATAAAGC 1195
 QY 1206 TAGCAATGGCTCTGTTATCAAAATGATGCTCCATTTGATCTACCAAGTATATAGATC 1265
 Db 1196 TAGCAATGGCTCTGTTATCAAAATGATGCTCCATTTGATCTACCAAGTATATAGATC 1255
 QY 1266 GGTACTCAGCCAAACAAAGCTTATGCTCTTTTATATCAGGTTCCCATGATGTGAAAAATGG 1325
 Db 1256 GGTACTCAGCCAAACAAAGCTTATGCTCTTTTATATCAGGTTCCCATGATGTGAAAAATGG 1315
 QY 1326 AGGTGAATTAATCAATCCACCATAGCCCGGCGGAGTCTCTCCCGCCCGCTCATCAG 1385
 Db 1316 AGGTGAATTAATCAATCCACCATAGCCCGGCGGAGTCTCTCCCGCCCGCTCATCAG 1375
 QY 1386 TCAGCGGTTTGTACCAACAAACAGGCTGCGGCGGCTTTATCGGACCAACAGCTTCCCTC 1445

Db 1376 TCAGCGGTTTGTCAACAAACAGGCTGCGCCAGGCTTTATCGGACCAACAGCTTCCCTC 1435
 QY 1446 TCACATGATAAAGATCCACCTCACTTAAATGGGACTCGACCATTTGAAAGACACGCCCAAG 1505
 Db 1436 TCACATGATAAAGATCCACCTCACTTAAATGGGACTCGACCATTTGAAAGACACGCCCAAG 1495
 QY 1506 CAGTTTCCATGTCGATCTCTAAACGGGAATTCAGTGTCAACAGGCTAGTCTCTGTTAAATGC 1565
 Db 1496 CAGTTTCCATGTCGATCTCTAAACGGGAATTCAGTGTCAACAGGCTAGTCTCTGTTAAATGC 1555
 QY 1566 TTACGCTTTCTGTCCAAAACTGGTCAAGTAAATAGTCTCTCAGTGATCCAGAACATCTTAA 1625
 Db 1556 TTACGCTTTCTGTCCAAAACTGGTCAAGTAAATAGTCTCTCAGTGATCCAGAACATCTTAA 1615
 QY 1626 GAAACAAAAAATTAACAATCAGTATTCACAAAGTTGCTGTTGCGCAGTGTCACTCTCA 1685
 Db 1616 GAAACAAAAAATTAACAATCAGTATTCACAAAGTTGCTGTTGCGCAGTGTCACTCTCA 1675
 QY 1686 ACCTAACCTTTCATAGTAAATTTCTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCAT 1745
 Db 1676 ACCTAACCTTTCATAGTAAATTTCTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCAT 1735
 QY 1746 TACCAATTTCTGAGTACAGTCTACCTCGAAACGATCTACGATGTCAAGTGTCTAGTAAAGT 1805
 Db 1736 TACCAATTTCTGAGTACAGTCTACCTCGAAACGATCTACGATGTCAAGTGTCTAGTAAAGT 1795
 QY 1806 AACAAACCGATCCCGGAGTGAATCTCTGCTCCAGCCGCTGATCAATGGCAAAATCCAA 1865
 Db 1796 AACAAACCGATCCCGGAGTGAATCTCTGCTCCAGCCGCTGATCAATGGCAAAATCCAA 1855
 QY 1866 GCTGAATCCAGCTGCTGCTGCTATGCGCGGAGTCTCTCTGAGGACTCTGACGAGGA 1925
 Db 1856 GCTGAATCCAGCTGCTGCTGCTATGCGCGGAGTCTCTCTGAGGACTCTGACGAGGA 1915
 QY 1926 GTCAAAAGGCTGCGCAAGAGGAATGGGATGTTGTTGATGTTGAGCTTCCCACTCTCCCGG 1985
 Db 1916 GTCAAAAGGCTGCGCAAGAGGAATGGGATGTTGTTGATGTTGAGCTTCCCACTCTCCCGG 1975
 QY 1986 CCAAGATGCGGAGATGAGGAGCCACTCGGACGAGCTTCAAGAACCCATGACCCCTAAA 2045
 Db 1976 CCAAGATGCGGAGATGAGGAGCCACTCGGACGAGCTTCAAGAACCCATGACCCCTAAA 2035
 QY 2046 CCGTGTCTTAATAGTGCAGACAGCAGTGAACCGAAAGAAACCGCTAGCGCTGATGG 2105
 Db 2036 CCGTGTCTTAATAGTGCAGACAGCAGTGAACCGAAAGAAACCGCTAGCGCTGATGG 2095
 QY 2106 TGCAGCTGCAAGGCGCAGCTGCTGCTGCTCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 2165
 Db 2096 TGCAGCTGCAAGGCGCAGCTGCTGCTGCTCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 2155
 QY 2166 TCTTCTGGAAGTTGATGCTGCTCTTCTGCTGCTCTCTCCAGAACCAAAATCTTTAGA 2225
 Db 2156 TCTTCTGGAAGTTGATGCTGCTCTTCTGCTGCTCTCTCCAGAACCAAAATCTTTAGA 2215
 QY 2226 GACCTTCAGGCTTAGCAACAAACTGAAAGCTCGAAGATGAATGATGATGATGATGATGATGATG 2285
 Db 2216 GACCTTCAGGCTTAGCAACAAACTGAAAGCTCGAAGATGAATGATGATGATGATGATGATGATG 2275
 QY 2286 AGAGAGGCGCTTCCAGAGACCGGAGCGGAGCTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 2345
 Db 2276 AGAGAGGCGCTTCCAGAGACCGGAGCGGAGCTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 2335
 QY 2346 ATCCCTGAGGAGTCAAGTGGCGGCGCAAGCTT 2378
 Db 2336 ATCCCTGAGGAGTCAAGTGGCGGCGCGGCT 2368

RESULT 7
 AX107851
 LOCUS AX107851
 DEFINITION Sequence 2 from Patent WO0123589.
 ACCESSION AX107851
 linear PAT 30-APR-2001

```
VERSION      AX107851.1  GI:13923242
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    1
AUTHORS      Kapeller-Libermann, R.
TITLE        23431, a novel human ubiquitin protease
JOURNAL      Patent: WO 0123589-A 2 05-APR-2001;
              Millennium Pharmaceuticals, Inc. (US)
FEATURES     Location/Qualifiers
              1..2347
               /organism="Homo sapiens"
               /mol_type="unassigned DNA"
               /db_xref="taxon:9606"
               /note="23431 Ubiquitin protease"
               61..>2347
                /note="unnamed protein product"
                /codon_start=1
                /protein_id="CAC37760.1"
                /db_xref="GI:13923243"
                /db_xref="REVIEW:13923243"
                /translation="MTIVDKASESSDPSAYQNGSSEAVSPDMDAGSASWGVSSL
NDVSHNLSLGPVGVAVYSSSVDPKSPQKQALGDGIAPKQVLPFPEKIKLK
WQTHRVGAGLQNLGTCFANAALQCLTYTPPLANTMLSHEHSKTHARFGCMWQ
AHIQALSNPGLVIGKPMFVINEMRRIRARFRGNOEDAEFLQYTVDMQKACLNGSN
KIDRHTQATLTCVQIFGGVLRVRKCLNCKGVSDPDPYLDITLLEIKAAQSVKALEQ
FYKPLQDGENSYKSKCKMVPASRFTIHRSSNVLTLULKRFANFTGGKIAQDVKY
PRYLDIRPSPNGPEPIVYLVAVLHVTGFNCHAGHYFCYIKASNLGYQWMDISV
TSDIRSVLSQQAIVLYFIRSHDVXNGELTHPSGQSPRPVLSQRVVVINKQAAPG
FTGPQLPSPHMIKNPFLHNGTGLKDTPPSSMSPNGNSVNRASPVNASVQNMNVN
RSVLPHPKPKQKTIITSHNKLVPQCSQPNLSNLENPTKVPSPSITINSVQST
SNASVMSVSKYTKPIPRSESCSQPMWKGKLSNLSVLPYGAESEDSDEESKGLGK
ENIGITIVSHSPGQDADEATPHELPQPMTLNGANSADSDPKENGLADPGASCQ
GQPALHSENPFAKANGLPKMPAPLLSLPDKILLETFLSNKLKXGSTDMSAPGAER
GPPEDRDABEPQGPSAESLSE"
```

Query Match 97.0%; Score 2339.2; DB 6; Length 2347;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2341; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 16 GCGGGCGGGCGCGCGAGGGGATGGAGCGAGCGCGCGAGCGGGTCAGAGTTGAACAATG 75
DB 4 GCGTCGCGGGCGCGAGGGGATGGAGCGAGCGCGCGAGCGGGTCAGAGTTGAACAATG 63
QY 76 ACCATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCCTATCAGAATCAGCGTGGC 135
DB 64 ACCATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCCTATCAGAATCAGCGTGGC 123
QY 136 AGCTCCGAGGAGTCTCACCTGGAGACATGGATGAGGTTCTGCGAGCTGGGGTCTGTG 195
DB 124 AGCTCCGAGGAGTCTCACCTGGAGACATGGATGAGGTTCTGCGAGCTGGGGTCTGTG 183
QY 196 TCTTCATTGAATGATGTGTCAAAATCACACACTTCTTTAGGACCAAGTACCTGCTGCTGTA 255
DB 184 TCTTCATTGAATGATGTGTCAAAATCACACACTTCTTTAGGACCAAGTACCTGCTGCTGTA 243
QY 256 GTTTATTCGAGTTGATCTGATCAATCAAAATCAAAACCATCACCAAAAGGATCAAGCC 315
DB 244 GTTTATTCGAGTTGATCTGATCAATCAAAATCAAAACCATCACCAAAAGGATCAAGCC 303
QY 316 CTAGTGATGGATCGCTCCCTCCAGAAAGTCTTTTCCCATCTGAGAAAGTTGTCTT 375
DB 304 CTAGTGATGGATCGCTCCCTCCAGAAAGTCTTTTCCCATCTGAGAAAGTTGTCTT 363
QY 376 AAGTGGCAACAACTCAVAGAGTTGGAGCTGGCTCCAGAAATTTGGGCAATACCTGTTTT 435
DB 364 AAGTGGCAACAACTCAVAGAGTTGGAGCTGGCTCCAGAAATTTGGGCAATACCTGTTTT 423
QY 436 GCCAATGACGACCTGACGTGTTTAACTTACACACCACTCTTGGCAATTAACATGCTATCA 495
```

QY	1576	GTCCAAAAC	CTGGTCA	AGTTAA	TAGTCC	TCAGTGAT	CCCGAAGACAT	CTCTAAGAAAA	CAAAAA	1635	
Db	1564	GTCCAAAAC	TGGTCA	AGTTAA	TAGTCC	TCAGTGAT	CCCGAAGACAT	CTCTAAGAAAA	CAAAAA	1623	
QY	1636	ATTACAAAT	CAGTATTC	CACAAAC	AGTTGG	CTGTTC	GCACAGTGT	CAGTCTCA	ACCTAACTTT	1695	
Db	1624	ATTACAAAT	CAGTATTC	CACAAAC	AGTTGG	CTGTTC	GCACAGTGT	CAGTCTCA	ACCTAACTTT	1683	
QY	1696	CATAGTAA	TTCTTTT	GGAGAA	CCCTAC	CAAGCCGGTTC	CTCTTAC	CAATACCAAT	TCTT	1755	
Db	1684	CATAGTAA	TTCTTTT	GGAGAA	CCCTAC	CAAGCCGGTTC	CTCTTAC	CAATACCAAT	TCTT	1743	
QY	1756	GCAGTACAG	TCTACCT	CGAACG	CATCTAC	GATGTCA	GTTCAGT	TTCTTAGT	AAAGTAA	CAAAAACCG	1815
Db	1744	GCAGTACAG	TCTACCT	CGAACG	CATCTAC	GATGTCA	GTTCAGT	TTCTTAGT	AAAGTAA	CAAAAACCG	1803
QY	1816	ATCCCCCG	CAGTGA	ATCTCT	CTCCAG	CCCGTGAT	GAATGG	CAAAAT	CCAAGCTGA	CTCTCC	1875
Db	1804	ATCCCCCG	CAGTGA	ATCTCT	CTCCAG	CCCGTGAT	GAATGG	CAAAAT	CCAAGCTGA	CTCTCC	1863
QY	1876	AGCGTGT	CTGTG	CGCTAT	TGCGCG	CGAGTCTCT	CTGAGG	ACTCTG	ACGAGAGT	CAAAAGGG	1935
Db	1864	AGCGTGT	CTGTG	CGCTAT	TGCGCG	CGAGTCTCT	CTGAGG	ACTCTG	ACGAGAGT	CAAAAGGG	1923
QY	1936	CTGGCAAG	GAGAA	TGGAT	TGGTAC	GAATGT	GAGCTT	CCCACTCT	CTCCGG	CCCAAGATGCC	1995
Db	1924	CTGGCAAG	GAGAA	TGGAT	TGGTAC	GAATGT	GAGCTT	CCCACTCT	CTCCGG	CCCAAGATGCC	1983
QY	1996	GAAGATGAG	GAGGCC	CACTCC	CGACG	AGCTTCA	AGAAC	CCCATG	ACCTTAA	ACGGTGCTAAT	2055
Db	1984	GAAGATGAG	GAGGCC	CACTCC	CGACG	AGCTTCA	AGAAC	CCCATG	ACCTTAA	ACGGTGCTAAT	2043
QY	2056	AGTCACAC	AGCAG	CACTG	ATGAC	CCGAA	AGAAA	CGGCTAG	CGCTGAT	GTGCGACCTGC	2115
Db	2044	AGTCACAC	AGCAG	CACTG	ATGAC	CCGAA	AGAAA	CGGCTAG	CGCTGAT	GTGCGACCTGC	2103
QY	2116	CAAGCCAG	AGCTGC	CTTGCA	CTCAG	AAAAAT	CCCTTTG	CTAAGG	CAAA	CGGTCTTCTGGA	2175
Db	2104	CAAGCCAG	AGCTGC	CTTGCA	CTCAG	AAAAAT	CCCTTTG	CTAAGG	CAAA	CGGTCTTCTGGA	2163
QY	2176	AAGTTGAT	AGCTGT	CTCTTTG	CTGTCT	CTCCAG	AGACAA	AAATCTT	TAGAGAC	CTTTCAGG	2235
Db	2164	AAGTTGAT	AGCTGT	CTCTTTG	CTGTCT	CTCCAG	AGACAA	AAATCTT	TAGAGAC	CTTTCAGG	2223
QY	2236	CTTAGCAC	ACAACT	GAAGG	CTCCAG	CGATGA	ATAGT	GCAC	CTGGAG	CAGAGAGGGC	2295
Db	2224	CTTAGCAC	ACAACT	GAAGG	CTCCAG	CGATGA	ATAGT	GCAC	CTGGAG	CAGAGAGGGC	2283
QY	2296	CCTCCGAG	GACCG	CGACG	CGCTCAG	CTTGG	ACGCC	CCGCG	CGGAAT	CCCTGGAG	2355
Db	2284	CCTCCGAG	GACCG	CGACG	CGCTCAG	CTTGG	ACGCC	CCGCG	CGGAAT	CCCTGGAG	2343
QY	2356	GAGC	2359								
Db	2344	GAGC	2347								

RESULT 8			
AX364931			
LOCUS	AX364931	1679 bp	DNA linear
DEFINITION	Sequence 82 from Patent WO0206315.		

VERSION AX364931.1 GI:18696821
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1
 REFERENCE Mintz, L., Freilich, S. and Bernstein, J.
 AUTHORS Novel nucleic acid and amino acid sequences
 TITLE Patent: WO 0206315-A 82 24-JAN-2002;
 JOURNAL

FEATURES	CompuGen Ltd. (IL)	Location/Qualifiers	1. .1679	69.5%;	Score 1675.6;	DB 6;	Length 1679;	0;
source			/organism="Homo sapiens"	99.8%;	Pred. No. 0;			
			/mol_type="unassigned DNA"					
			/db_xref="taxon:9606"					
ORIGIN								
	Query Match							
	Best Local Similarity							
	Matches 1675; Conservative							
Qy	694	GTATGCGAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCACC	753					
Db	1	GCTATGCGAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCACC	60					
Qy	754	ACTCTTGTTTGTCCAGATATTTGGAGGATACCTAAGATCTTAGAGTCAAATGTTTAAATTGC	813					
Db	61	ACTCTTGTTTGTCCAGATATTTGGNGGATACCTAAGATCTTAGAGTCAAATGTTTAAATTGC	120					
Qy	814	AAGGGCGTTTCAGTACTTTTGATTCATATCTTTGATATAAATTTGGAGATAAAGGCTGCT	873					
Db	121	AAGGGCGTTTCAGTACTTTTGATTCATATCTTTGATATAAATTTGGAGATAAAGGCTGCT	180					
Qy	874	CAGAGTGTCAACNAGCANTGGAGCAGTTTGTGAAGCCGGAACAGCTTCATGGAGAAAC	933					
Db	181	CAGAGTGTCAACNAGCANTGGAGCAGTTTGTGAAGCCGGAACAGCTTCATGGAGAAAC	240					
Qy	934	TCGTACAAGTGCAGCAAGTGTAAAGATGGTTTCCAGCTTCAAAGAGGTTCACTATCCAT	993					
Db	241	TCGTACAAGTGCAGCAAGTGTAAAGATGGTTTCCAGCTTCAAAGAGGTTCACTATCCAT	300					
Qy	994	AGATCCCTCTAATGTTCTTACATTTCTCTGAAACGGTTTTTGCAAAATTTTACCGGTGGA	1053					
Db	301	AGATCCCTCTAATGTTCTTACATTTCTCTGANACGTTTTTGCAAAATTTTACCGGTGGA	360					
Qy	1054	ATTGCTTAAGATGTGAATAACCTCGAGTACTTGTATATTCGGCCATATATGTTCTCAACC	1113					
Db	361	ATTGCTTAAGATGTGAATAACCTCGAGTACTTGTATATTCGGCCATATATGTTCTCAACC	420					
Qy	1114	AACGGAGAGCAATTTGTCTACGTCTTGATGTCAGTCTGGTCCACACTGGTTTTAAATGCG	1173					
Db	421	AACGGAGAGCAATTTGTCTACGTCTTGATGTCAGTCTGGTCCACACTGGTTTTAAATGCG	480					
Qy	1174	CATGCTGGCCATTAATTTCTGCTCAATAAAGCTAGCAATGGGCTCTGGTATCAATGAAT	1233					
Db	481	CATGCTGGCCATTAATTTCTGCTCAATAAAGCTAGCAATGGGCTCTGGTATCAATGAAT	540					
Qy	1234	GACTTCCATTTGATCTACAGTGATATTTAGATCGGTACTAGCCAAACAAAGCTATGTGCTC	1293					
Db	541	GACTTCCATTTGATCTACAGTGATATTTAGATCGGTACTAGCCAAACAAAGCTATGTGCTC	600					
Qy	1294	TTTTATATCAGGTCCCATGATGTGAAATATGGAGGTGAATTAATCATCCACCCCATAGC	1353					
Db	601	TTTTATATCAGGTCCCATGATGTGAAATATGGAGGTGAATTAATCATCCACCCCATAGC	660					
Qy	1354	CCCGGCCAGTCTCTCCCGGCCCGCTCATCAGTCCAGCGGTTGTCAACCAACAAAGGCT	1413					
Db	661	CCCGGCCAGTCTCTCTCCCGGCCCGCTCATCAGTCCAGCGGTTGTCAACCAACAAAGGCT	720					
Qy	1414	GGCGCCAGCTTTATCGGACCAACAGTTCCTCTCACATGATTAAGAAATCCACCTCACTTA	1473					
Db	721	GGCGCCAGCTTTATCGGACCAACAGTTCCTCTCACATGATTAAGAAATCCACCTCACTTA	780					
Qy	1474	AATGGGACTGACCAATTTGAAAGACACGCCAAGAGTTCATGTGAGTCTTAACGGGAAT	1533					
Db	781	AATGGGACTGACCAATTTGAAAGACACGCCAAGAGTTCATGTGAGTCTTAACGGGAAT	840					
Qy	1534	TCCAGTGTCAACAGGGCTAGTCTGTTAATGCTTCAGCTTCTGTCCAAAACCTGGTCAGTT	1593					
Db	841	TCCAGTGTCAACAGGGCTAGTCTGTTAATGCTTCAGCTTCTGTCCAAAACCTGGTCAGTT	900					
Qy	1594	AATAGGTCTTCAGTGTATCCAGACATCTCTTAAGAAACAAAAATTAACAATCAGTATTTAC	1653					

[illegible]

```
Db 901 AATAGTCTCTCAGTATCCAGAACATCTCTAAGAAACAAATTAACAATCAGTATTCAC 960
QY 1654 AACAGTGGCTCTGCGCAGTGTGAGTCTCAACCTTATAGTAAATCTTTGGAG 1713
Db 961 AACAGTGGCTCTGCGCAGTGTGAGTCTCAACCTTATAGTAAATCTTTGGAG 1020
QY 1714 AACCTTACCAAGCCGCTTCCCTCTTACCAATTAACAATTAACAATTAACAATTAACAAT 1773
Db 1021 AACCTTACCAAGCCGCTTCCCTCTTACCAATTAACAATTAACAATTAACAATTAACAAT 1080
QY 1774 AAGCATCTACGATGTGAGTCTTCTAGTAAAGTAAACAAACCCGCGAGTGAATCC 1833
Db 1081 AAGCATCTACGATGTGAGTCTTCTAGTAAAGTAAACAAACCCGCGAGTGAATCC 1140
QY 1834 TGCTCCACGCGTGTGATGCAATCAAGTCAACCTCAAGTCAAGTCAAGTCAAGTCAAGT 1893
Db 1141 TGCTCCACGCGTGTGATGCAATCAAGTCAACCTCAAGTCAAGTCAAGTCAAGTCAAGT 1200
QY 1894 GCGCGAGTCTCTGAGGACTCTGACGAGGAGTCAAGGCGTCAAGGCGTCAAGGCGTCAAGG 1953
Db 1201 GCGCGAGTCTCTGAGGACTCTGACGAGGAGTCAAGGCGTCAAGGCGTCAAGGCGTCAAGG 1260
QY 1954 ATTGTGATGATGTGAGTCTTCCACTCTCCCGGCAAGATGCGGAGTCAAGGCGTCAAGG 2013
Db 1261 ATTGTGATGATGTGAGTCTTCCACTCTCCCGGCAAGATGCGGAGTCAAGGCGTCAAGG 1320
QY 2014 CGCAGGAGTCTCAAGAACCCATGACCTTAACCGTGTCTAATGTGACAGACGACAGT 2073
Db 1321 CGCAGGAGTCTCAAGAACCCATGACCTTAACCGTGTCTAATGTGACAGACGACAGT 1380
QY 2074 GACCCGAAAGAAACGGCTAGCGCTGATGTGCGCAGTGCACAGGCGCAGGCGTCCCTG 2133
Db 1381 GACCCGAAAGAAACGGCTAGCGCTGATGTGCGCAGTGCACAGGCGCAGGCGTCCCTG 1440
QY 2134 CACTCAGAAATCCCTTGTGTAAGCAACCGTCTTCTTGGAAAGTGTGATGCTGCTCT 2193
Db 1441 CACTCAGAAATCCCTTGTGTAAGCAACCGTCTTCTTGGAAAGTGTGATGCTGCTCT 1500
QY 2194 TTGCTGTCTCTCCAGACAGCAAAATCTAGACGCTTCAGGCTTAGCAAACTGAAA 2253
Db 1501 TTGCTGTCTCTCCAGACAGCAAAATCTAGACGCTTCAGGCTTAGCAAACTGAAA 1560
QY 2254 GGCTCGACGATGAATGATGTCACCTTGAGCAGAGAGGGGCTCCCGAGGACGCGAC 2313
Db 1561 GGCTCGACGATGAATGATGTCACCTTGAGCAGAGAGGGGCTCCCGAGGACGCGAC 1620
QY 2314 GCGAGCTCTAGCTGCGAGCCCGCGCGAATCCCTGGAGGACCGAGATGCGGCGCG 2372
Db 1621 GCGAGCTCTAGCTGCGAGCCCGCGCGAATCCCTGGAGGACCGAGATGCGGCGCG 1679
```

```
RESULT 9
AC004895/c
LOCUS AC004895 152927 bp DNA linear PRI 30-SEP-2000
DEFINITION Homo sapiens PAC clone RP4-810E6 from 7, complete sequence.
ACCESSION AC004895
VERSION AC004895.2 GI:4926908
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 152927)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 152927)
Cordes, M. and Gibson, A.
The sequence of Homo sapiens PAC clone RP4-810E6
Unpublished
```

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

3 (bases 1 to 152927)
Waterston, R.H.
Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

4 (bases 1 to 152927)
Waterston, R.H.
Direct Submission
Submitted (29-MAY-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

5 (bases 1 to 152927)
Waterston, R.
Direct Submission
Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 29, 1999 this sequence version replaced gi:3450909.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_DJ0810E06

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP1-42M2. Actual start of this clone is at base position 1 of RP4-810E6; actual end is at 152927 of RP4-810E6.

There is a questionable number of A's from 81574-81601. The consensus reflects 27 A's but could be as many as 29.

FEATURES
source
1. .152927
/Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"

```
/map="7"
/clone="RP4-810E6"
/clone_lib="RPOI-4"
329. .464
/rpt_family="L1"
repeat_region
466. .560
/rpt_family="MIR"
repeat_region
649. .833
/rpt_family="Alu"
repeat_region
1269. .1585
/rpt_family="Alu"
repeat_region
1586. .1745
/rpt_family="L1"
repeat_region
1813. .1974
/rpt_family="Alu"
repeat_region
2887. .2970
/rpt_family="Alu"
repeat_region
2977. .3093
/rpt_family="L1"
repeat_region
3094. .3392
/rpt_family="Alu"
repeat_region
3393. .3448
/rpt_family="L1"
repeat_region
3449. .3751
/rpt_family="Alu"
repeat_region
3761. .4060
/rpt_family="Alu"
repeat_region
4061. .4162
/rpt_family="L1"
repeat_region
4163. .4485
/rpt_family="Alu"
repeat_region
4486. .4573
/rpt_family="L1"
repeat_region
4574. .4689
/rpt_family="Alu"
repeat_region
4691. .4988
/rpt_family="L1"
repeat_region
4969. .4988
/rpt_family=" (TAAA)n"
repeat_region
4992. .5112
/rpt_family="Alu"
repeat_region
5121. .5179
/rpt_family="L1"
repeat_region
5180. .5486
/rpt_family="Alu"
repeat_region
5487. .5736
/rpt_family="L1"
repeat_region
5912. .5965
/rpt_family="A-rich"
repeat_region
6214. .6526
/rpt_family="Alu"
repeat_region
6544. .6601
/rpt_family="L2"
repeat_region
6673. .6984
/rpt_family="Alu"
repeat_region
7523. .7625
/rpt_family="MIR"
repeat_region
7914. .8236
/rpt_family="Alu"
misc_feature
8558. .8624
/notes="match to EST M79246 (NID:g273559)"
9165. .9231
/notes="match to EST M79246 (NID:g273559)"
10016. .10318
/rpt_family="Alu"
repeat_region
10529. .10564
/rpt_family="Alu"
repeat_region
10565. .10867
/rpt_family="Alu"
repeat_region
10868. .10977
/rpt_family="Alu"
repeat_region
11406. .11705
/rpt_family="Alu"
repeat_region

11922. .12231
/rpt_family="Alu"
repeat_region
12233. .12567
/rpt_family="MaLR"
repeat_region
13482. .13754
/rpt_family="Alu"
repeat_region
13771. .13925
/rpt_family="Alu"
repeat_region
13932. .14240
/rpt_family="Alu"
repeat_region
14520. .14937
/rpt_family="Other"
repeat_region
15405. .15431
/rpt_family="AT-rich"
repeat_region
15611. .15852
/rpt_family="MaLR"
repeat_region
16336. .16460
/rpt_family=" (CA)n"
repeat_region
16480. .16693
/rpt_family=" (CA)n"
repeat_region
16793. .17187
/rpt_family="L1"
repeat_region
17675. .17877
/rpt_family="MaLR"
misc_feature
18340. .18452
/notes="match to EST M79246 (NID:g273559)"
18400. .18458
/notes="similar to EST AA284549 (NID:g1927478) zt24f01.r1"
19073. .19720
/rpt_family="L1"
repeat_region
19721. .20026
/rpt_family="Alu"
repeat_region
20027. .20435
/rpt_family="L1"
repeat_region
20436. .20735
repeat_region

Query Match 30.1%; Score 727.2; DB 9; Length 152927;
Best Local Similarity 98.3%; Pred. No. 7.6e-197;
Matches 735; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1435 CAGCTTCCTTCACATGATAAGAAATCCACCTCCTTAATAGGACCTGGACCATGAAA 1494
Db 46721 CAGGTGGCTCTGTACTTCCAGATCCACCTCCTTAATAGGACCTGGACCATGAAA 46662
Qy 1495 GACACGCCAAGCACTTCATGTCGAGTCCTAAACGGGAATTCAGTGTCAACAGGGCTAGT 1554
Db 46661 GACACGCCAAGCACTTCATGTCGAGTCCTAAACGGGAATTCAGTGTCAACAGGGCTAGT 46602
Qy 1555 CCTGTTAATGCTTCAGCTTCGTCTCAAAAATGTTGAGTATAGTCTCAGTATCCCA 1614
Db 46601 CCTGTTAATGCTTCAGCTTCGTCTCAAAAATGTTGAGTATAGTCTCAGTATCCCA 46542
Qy 1615 GAACATCCTAAGAACCAAAAATACATCATGATGATTCACAAAGATTCCTGTTCCGCGAG 1674
Db 46541 GAACATCCTAAGAACCAAAAATACATCATGATGATTCACAAAGATTCCTGTTCCGCGAG 46482
Qy 1675 TGTGAGTCTCAACCTTAACCTTCATAGTAATTCCTTGGAGAACCTTACCAAGCCGCTTCCC 1734
Db 46481 TGTGAGTCTCAACCTTAACCTTCATAGTAATTCCTTGGAGAACCTTACCAAGCCGCTTCCC 46422
Qy 1735 TCTTCTACCATTAACAAATCTGAGTACAGTCTACCTGAGAGGATCTACGATGTCAGTT 1794
Db 46421 TCTTCTACCATTAACAAATCTGAGTACAGTCTACCTGAGAGGATCTACGATGTCAGTT 46362
Qy 1795 TCTAGTAAAGTAAACAAACCGATCCCGCAGTGAATCTGCTCCAGCCGCTGATGAAT 1854
Db 46361 TCTAGTAAAGTAAACAAACCGATCCCGCAGTGAATCTGCTCCAGCCGCTGATGAAT 46302
Qy 1855 GGCNAATCCAAAGCTGAATCCAGCGTGTGTTGCCCTATGGCGCGAGTCTCTTGAGGAC 1914
Db 46301 GGCNAATCCAAAGCTGAATCCAGCGTGTGTTGCCCTATGGCGCGAGTCTCTTGAGGAC 46242
Qy 1915 TCTGACGAGGAGTCAAGGGGCTGGCAAGGAGATGGATTTGGTACGATTTGTGAGCTCC 1974
```


Df	46241	TCTGACGAGAGTCAAAAGGGCGTCGGCCAGGAGAATGGGATTGTGTAAGTTGTGAGCTCC	46182
Qy	1975	CACCTCTCCCGGCCAAGATGCGGAAGATGAGAGGCCACTCGCACGAGCTTCAAGAACCC	2034
Df	46181	CACCTCTCCCGGCCAAGATGCGGAAGATGAGAGGCCACTCGCACGAGCTTCAAGAACCC	46122
Qy	2035	ATGACCCTTAACGGTGCTTAATAGTCAGACAGCAGCAGTAGTGACCCGAAAGAAAACGGCTA	2094
Df	46121	ATGACCCTTAACGGTGCTTAATAGTCAGACAGCAGCAGTAGTGACCCGAAAGAAAACGGCTA	46062
Qy	2095	GCGCCTGATGGTGCAGCTGCCAAGCGCAGCCTGCCCTGCACTCAGAAAAATCCCTTTGCT	2154
Df	46061	GCGCCTGATGGTGCAGCTGCCAAGCGCAGCCTGCCCTGCACTCAGAAAAATCCCTTTGCT	46002
Qy	2155	AAGGCAACCGGTCTTCTCCTGGAAAGTTGA	2182
Df	46001	AAGGCAACCGGTCTTCTCCTGGAAAGTTGA	45974
RESULT 10			
AC069393		154791 bp DNA linear HTG 26-OCT-2000	
LOCUS	Homo sapiens chromosome 7 clone RP11-585K18, WORKING DRAFT		
DEFINITION	SEQUENCE, 14 unordered pieces.		
ACCESSION	AC069393		
VERSION	AC069393.8 GI:11024947		
KEYWORDS	HTG; HTGS PHASE1; HTGS_DRAFT.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 154791)		
JOURNAL	Waterston,R.H.		
REFERENCE	The sequence of Homo sapiens clone		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 154791)		
JOURNAL	Waterston,R.H.		
COMMENT	Direct Submission Submitted (28-MAY-2000) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA On Oct 26, 2000 this sequence version replaced gi:10946568. ----- Genome Center ----- Center: Washington University Genome Sequencing Center Center code: WUGSC Web site:http://genome.wustl.edu/gsc/index.shtml ----- Project Information ----- Center project name: H NH0585K18 ----- Summary Statistics ----- Sequencing vector: M13, 71% Chemistry: Dye-primer ET; 71% of reads Assembly program: Phrap; version 0.990319 Consensus quality: 143345 bases at least Q40 Consensus quality: 146572 bases at least Q30 Consensus quality: 148386 bases at least Q20 Insert size: 164000; agarose-fp Insert size: 153491; sum-of-contigs Quality coverage: 4.65 in Q20 bases; agarose-fp Quality coverage: 5.04 in Q20 bases; sum-of-contigs * NOTE: This is a 'working draft' sequence. It currently consists of 14 contigs. The true order of the pieces * is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. *		
	1	5209: contig of 5209 bp in length	

Db	129179	CAGTSCGCTCTGCTTACTTCCAGAAATCACCTCACTTAATATGGAGCTGGACCATTTGAA	129238
QY	1495	GACAGCCCAAGCAGTTCCATGTCGAGTCCTAACGGGAATTCAGGTGTCACAGGGCTAGT	1554
Db	129239	GACAGCCCAAGCAGTTCCATGTCGAGTCCTAACGGGAATTCAGGTGTCACAGGGCTAGT	129298
QY	1555	CCTGTTAATGCTTCAGCTTCTGTCAAAACCTGGTCAGTTAATAGTTCCTCAGTGATCCCA	1614
Db	129299	CCTGTTAATGCTTCAGCTTCTGTCAAAACCTGGTCAGTTAATAGTTCCTCAGTGATCCCA	129358
QY	1615	GAACATCTTAAGAAACAAAATTAACATCAGTATTCAACAAGTTGCTGTTCGCCAG	1674
Db	129359	GAACATCTTAAGAAACAAAATTAACATCAGTATTCAACAAGTTGCTGTTCGCCAG	129418
QY	1675	TGTGAGTCTCAACCTCACTTATAGTAATCTTTGGAGAACCTTACCAAGCCCGTTGCC	1734
Db	129419	TGTGAGTCTCAACCTCACTTATAGTAATCTTTGGAGAACCTTACCAAGCCCGTTGCC	129478
QY	1735	TCCTTACCATTACCAATTCGAGTACGATCTACCTCGAACGCATCTACGATTCAGTT	1794
Db	129479	TCCTTACCATTACCAATTCGAGTACGATCTACCTCGAACGCATCTACGATTCAGTT	129538
QY	1795	TCTAGTAAAGTACAAAACCGATCCCCCGCAGTGAATCTCTGCTCCAGCCCGTGATGAAT	1854
Db	129539	TCTAGTAAAGTACAAAACCGATCCCCCGCAGTGAATCTCTGCTCCAGCCCGTGATGAAT	129598
QY	1855	GGCAAAATCCAAAGCTGAACTCCAGCGTCTGGTGCCCTATGCGCCGAGTCTCTGAGGAC	1914
Db	129599	GGCAAAATCCAAAGCTGAACTCCAGCGTCTGGTGCCCTATGCGCCGAGTCTCTGAGGAC	129658
QY	1915	TTGACGAGAGTCAAAAGGGCTGGGCAAGAGAAATGGGATTTGGTACGATTTGAGTCC	1974
Db	129659	TTGACGAGAGTCAAAAGGGCTGGGCAAGAGAAATGGGATTTGGTACGATTTGAGTCC	129718
QY	1975	CATCTCCCGGCCAAGATCGGAGATGAGAGGCCACTCGCACGAGCTTCAAGAACCC	2034
Db	129719	CATCTCCCGGCCAAGATCGGAGATGAGAGGCCACTCGCACGAGCTTCAAGAACCC	129778
QY	2035	ATGACCTTAAACGGTGTCTAATAGTGCAGACAGCGACAGTGCACCGAAAGAAACGGGCTA	2094
Db	129779	ATGACCTTAAACGGTGTCTAATAGTGCAGACAGCGACAGTGCACCGAAAGAAACGGGCTA	129838
QY	2095	GCGCTGATGTCAGTGCAGTGCACAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2154
Db	129839	GCGCTGATGTCAGTGCAGTGCACAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	129898
QY	2155	AAGCAACCGTCTTCTGGAAAGTTGA	2182
Db	129899	AAGCAACCGTCTTCTGGAAAGTTGA	129926
RESULT 11			
AC146146/c	217022 bp DNA linear HTG 01-AUG-2003		
LOCUS	Pan troglodytes chromosome UNK clone RP43-2G9, WORKING DRAFT		
DEFINITION	SEQUENCE, 26 unordered pieces.		
ACCESSION	AC146146		
VERSION	AC146146.1 GI:33387100		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	Pan troglodytes (chimpanzee)		
ORGANISM	Pan troglodytes		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.		
TITLE	1 (bases 1 to 217022)		
JOURNAL	Wilson, R.K.		
REFERENCE	The sequence of Pan troglodytes clone		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 217022)		
JOURNAL	Wilson, R.K.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444		
TITLE	Forest Park Parkway, St. Louis, MO 63108, USA		
JOURNAL			
COMMENT			

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: C_PT002G09
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 204681 bases at least Q40
Consensus quality: 206693 bases at least Q30
Consensus quality: 208289 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1098: contig of 1098 bp in length
* 1099: gap of unknown length
* 1199: contig of 1784 bp in length
* 2982: gap of unknown length
* 3083: contig of 2098 bp in length
* 5180: gap of unknown length
* 5181: contig of 1872 bp in length
* 7152: gap of unknown length
* 7252: contig of 2197 bp in length
* 9449: gap of unknown length
* 9549: contig of 1394 bp in length
* 10943: gap of unknown length
* 11043: contig of 2769 bp in length
* 13812: gap of unknown length
* 13912: contig of 4441 bp in length
* 13933: gap of unknown length
* 13953: gap of unknown length
* 14553: contig of 3222 bp in length
* 21675: gap of unknown length
* 21775: contig of 2532 bp in length
* 24307: gap of unknown length
* 24407: contig of 3805 bp in length
* 28212: gap of unknown length
* 28312: contig of 6382 bp in length
* 34694: gap of unknown length
* 34794: contig of 3484 bp in length
* 38278: gap of unknown length
* 38279: contig of 6494 bp in length
* 38379: gap of unknown length
* 44972: contig of 6550 bp in length
* 51522: gap of unknown length
* 51622: contig of 7708 bp in length
* 51623: gap of unknown length
* 59430: contig of 8255 bp in length
* 59431: gap of unknown length
* 67685: gap of unknown length
* 67686: contig of 8519 bp in length
* 76304: gap of unknown length
* 76305: contig of 15280 bp in length
* 91685: gap of unknown length
* 91785: contig of 11997 bp in length
* 103785: gap of unknown length
* 103882: contig of 12102 bp in length
* 115983: gap of unknown length
* 116083: contig of 15745 bp in length
* 131828: gap of unknown length
* 131829: contig of 16846 bp in length
* 131929: gap of unknown length
* 148774: contig of 19451 bp in length
* 148775: gap of unknown length
* 168325: contig of 19451 bp in length

168326	169425:	gap of unknown length
168426	185595:	contig of 17170 bp in length
185596	185695:	gap of unknown length
185696	217022:	contig of 31327 bp in length.

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10

```

1. 210922
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="UNK"
/clone="RP43-2G9"
1. 1098

```

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

misc feature

OFFICE OF PUBLIC DEFENSE

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

11

1

1

1

1

STANDARD - 2000

ATTN: _____

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

ORIGIN

	Query Match	29.8%	Score 714.4	DB 2	Length 217022	
	Best Local Similarity	37.2%	Pred. No. 3.9e-193			
	Matches 727	Conservative	0	Mismatches 21	Indels 0	Gaps 0
QY	1435	CAGCTTCCCTCTCCATGATAAAGATCCACCTCACTTAATGGGACTGCACCATGAAA	1494			
Db	150493	CAGGTGGCTTCTGCTTACTTCCAGATCCACCTCACTTAATGGGACTGCACCATGAAA	1504			

[illegible]

Db
150493 CAGGTGCGCTCTGGCTTACCTTCACGATCCACCTCACTTTGCCCTCCTGCCT

QY	1495	GRACGCCAGCAGTTCAGATGTCAGTCTTAACGGGAATTCAGTGTCAACAGGGCTAGT	1554
Db	150433	GRACGCCAAGCAGTTCAGATGTCAGTCTTAACGGGAATTCAGTGTCAACAGGGCTAGT	150374
QY	1555	CCTGTTAATGCTTCAGCTTCTGTCCAAACTCGTCAAGTTAATAAGTTCCTCAGTGATCCCA	1614
Db	150373	CCTGTTAATGCTTCAGCTTCTGTCCAAACTCGTCAAGTTAATAAGTTCCTCAGTGATCCCA	150314
QY	1615	GAACATCTTAAGAAACAAAATAATCAATCAGTATTCAACAAGTGTGCTGTTCGCCAG	1674
Db	150313	GAACATCTTAAGAAACAAAATAATCAATCAGTATTCAACAAGTGTGCTGTTCGCCAG	150254
QY	1675	TGTGAGTCTCAACCTTAACCTTCATAGTAATCTTTGGAGAACCTTACAGAGCCCGTTCC	1734
Db	150253	TGTGAGTCTCAACCTTAACCTTCATAGTAATCTTTGGAGAACCTTACAGAGCCCGTTCC	150194
QY	1735	TCTTCTACAAATPACAAATTCGAGTACAGTCTACCTCGAACGCATCTACGATGTCAGTT	1794
Db	150193	TCTTCTACAAATPACAAATTCGAGTACAGTCTACCTCGAACGCATCTACGATGTCAGTT	150134
QY	1795	TCTAGTAAAGTAACAAACCCGATCCCGCAGTGAATCCTGCTCCAGCCCGTGATGAAT	1854
Db	150133	TCTAGTAAAGTAACAAACCCGATCCCGCAGTGAATCCTGCTCCAGCCCGTGATGAAT	150074
QY	1855	GGCAAAATCCAAGCTGAATCTCAGCGTGTGTGTGCCCTATGGCGCGAGTCTCTCTCAGAC	1914
Db	150073	GGCAAAATCCAAGCTGAATCTCAGCGTGTGTGTGCCCTATGGCGCGAGTCTCTCTCAGAC	150014
QY	1915	TCTGACGAGGAGTCAAGGGGCTGGCGAAGAGAAATGGGATTTGTCAGATTGTGAGCTCC	1974
Db	150013	TCTGACGAGGAGTCAAGGGGCTGGCGAAGAGAAATGGGATTTGTCAGATTGTGAGCTCC	149954
QY	1975	CACCTCTCCCGGCCAAGATGCCGAAGATGAGCAGGCACCTCCGACAGAGCTTCAAGAACCC	2034
Db	149953	CACCTCTCCCGGCCAAGATGCCGAAGATGAGCAGGCACCTCCGACAGAGCTTCAAGAACCC	149894
QY	2035	ATGACCTTAACGGTGTCTAATAGTGTGAGACAGGCACAGTACCCGGAAGAAACGGGCTA	2094
Db	149893	ATGACCTTAACGGTGTCTAATAGTGTGAGACAGTACAGTACCCGGAAGAAATGGGCTG	149834
QY	2095	GGCGCTGATGTGTGCAGCTGCCAAGGCCAGCTCGCCTGCATCAGAAATCCCTTTGCT	2154
Db	149833	GGCGCTGATGTGTGCAGCTGCCAAGGCCAGCTCGCCTGCATCAGAAATCCCTTTGCT	149774
QY	2155	AAGCAAAACGGTCTTCTCGAAAAGTTGA	2182
Db	149773	AAGCAAAACGGTCTTCTCGAAAAGTTGA	149746

RESULT 12

BC057482	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
----------	-------	------------	-----------	---------	----------	--------	----------

SOURCE
ORGANISMS

ORGANISMS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J.G., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldi, M.F., Casavant, T.L.,
Schetzl, P.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carroll, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullaby, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Wortley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulvyk, S.W.

Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettenan, M., Madan, A., Rodrigues, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalhus, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A. Curation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257
12477932
2 (bases 1 to 3075)
Strausberg, R.
Direct Submission
Submitted (03-SEP-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue procurement: Dr. Sumio Sugano
cDNA Library Preparation: Dr. Sumio Sugano
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mdc@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 122 Row: 0 Column: 8
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Genomescan gene prediction, Similarity but not identity to protein.

FEATURES

1..3075
Location/Qualifiers
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:5915397"
/tissue_type="Whole body, adult male"
/clone_lib="Sugano SJD adult male"
/lab_host="DH10B"
/note="Vector: pME18S-FL3"
120..>3075
/codon_start=1
/product="Unknown (protein for IMAGE:5915397)"
/protein_id="AAH57482.1"
/db_xref="GI:3478578"
/translation="MESSSSVSPSATELSRAKTVCMAPAIAGATVNGSNNTALQVE
RPEQVGTSGDGLAKPKVLFSPERLCNKGHRIGAGLHNLGNTCFNLSTQLCT
YTAFLANYLTFRESHKTCHPEFCMMCTQNHIIQVFANSNVKIPSLNELRIK
HFRGSEDAHEFLRYTVDMOKSLPGNKLDRQTAFTVHQIFGGLRSRVKCLNC
KAVSDTFPYLDISLEIKTAOTLSKAFQFVKPEQLDGNAYKCKKKOMVASKRFT
VHRSNNVLSIKRATFNFGKIKTVRYAEHLDRPFMSQSHGEPQIYALVAVLHS
GFSCHAGHYCYIKASNGQWYONDSSVSLSDIRVLNQAYLLFYIRSPDVKNQGF
SOMNHPQSSPRSPISPLKNGPQVTTTSFIPOLPMLKNSVYVNGNSKKEVHG
SKPNRNFQGMTKIGPSHTSSASASSSLPVRSMGIPDSSKRPXLTFLIGKQVPA
QTQSPNCSLSASHSSTSTSTSTSGFPQVQVQVTHTGKAFLLVPGQSESESD
QEAQVLENGSACTYLAIPKTINGVQKDAQSSSLPHHTSKTNGSNFVCEGSD
SHVCENGSPDHTSQNGLSRANGFNHNDKMTDPTSSQISDLSCSEMELNLPKHN
ESRSPNPSADHPCSPSNKELNLPSSSSQVAOIQSSPKESLVLKTTDTDLHET
ASKVELTORHATHATGTSIDLVNNVSLVSHGHTSPHKLKVVYRCSEKAPKS
HSDKFGSDGGRNDRNGSDRPHFSSTPKDRERYRHYRHDSDRSRSGYSDSR
RSTGRERYRDRDLERHWRFRSHHREHHYFQRHRRDNRDRFRGSDSYRPSGYH
NRNGYSNHSRGMEEAHRATHVTNGSKRPSFSPHSVPLPRHHRKRSKPSVDARESS
DECKAKGKKKKK"
369..1265

misc_feature

/note="COG5077; Region: COG5077, Ubiquitin
carboxyl-terminal hydrolase [Posttranslational
modification, protein turnover, chaperones]"
/db_xref="CDD:COG5077"

ORIGIN

	Query Match	25.4%	Score 613.4	DB 5	Length 3075
	Best Local Similarity	66.1%	Pred. No. 2.1e-164		
	Matches 951	Conservative	Mismatches 476	Indels 12	Gaps 4
QY	67	TGAACAATGACCATAGTTGACAAAGCTTCTGAATCTTCAGACC---CATCAGCCTATCAG	123		
DB	21	TGATAAATGACCATAGTTGACAAACCTCCAGAGAAACCTGACCTGGAGTCAGTGGCGTGC	80		
QY	124	AATCAGCCTGGCAGCTCCGAGGCAGTCTCACCTGGAGACATGATGATGATGATGATGATG	183		
DB	81	AAGCACTCCAGCTTCATCACCATTCTCTCCAGAGACATGGAGAGCAGTAGTTCCGAGC	140		
QY	184	TGGGGTGTCTGTCTTCATTGAATGATGTGTGAATATCAC---ACATTTCCTTTAGGACCA	240		
DB	141	TGGAGTGTGGCCCTTCTGCACTGAGCTTCAAGGGCCCAAGACGGTCTGTATGGCCCCG	200		
QY	241	GTACCTGGTGTCTGATGTTTATTCGAGTTTCAT---CTGTACCTGATAAATCAAAACCATCA	297		
DB	201	GCCATTGGAGCGCATCTGTGAATGGTATGATAACACAGCACTCCAGTCGAAAGCCACGA	260		
QY	298	CCACAAAAGATCAAGCCCTAGGTGATGCGATCGCTCTCCACAGAAAAGTTCTTTTCCCA	357		
DB	261	GAGCAAGTGGGAGACAGTGGGATGGATGCTCTCCCTCAAAGGTCCTGTTTCC	320		
QY	358	TCTGAGAAGATTTCTTTAAGTGGCAACAACTCATATAGTGTGGAGCTGGGCTCCAGAT	417		
DB	321	CCAGAGGGGCTCTGCTTGAAGTGAACCAAGGCCACCGCATTTGGAGCAGGTCTCCATAAC	380		
QY	418	TTGGCAATACCTGTTTGGCAATGAGCACTGCAGTGTGTTTAACTTACACACCACTCTT	477		
DB	381	CTGGGCAACACATGTTTCTCGAACTCTACCTCGAGTGTCTGACCTACATCTGCTCCCTT	440		
QY	478	GCCAAATTACATGCTATCACATGAACACTCCAAACCATGTCTATGAGAGGCTTTTGTATG	537		
DB	441	GCCAACTACATGCTGACCAAGAGCATTTCCAAACATGTCTATGAGCCCGGTTTGTATG	500		
QY	538	ATGTGTAATGCAAGCACATATTAACCCAGGCACTCAGTATCTCTGGGAGGTTTATTA	597		
DB	501	ATGTGTACCATGCAAAATCACAATCATCCAAAGTGTGTCCTGCAACTCTGAAATGTCA	560		
QY	598	CCATGTTTGTCTCAATGATGATGCGCGGTATAGCTAGGCACTCCCGTTTGGAAACCAA	657		
DB	561	CCCATCAGTGTGCTTAATGAATGAAAAGGATTTGAAAGCATTTTAGATTTGGAAGTCAA	620		
QY	658	GAAGATGCCATGAATTCCTTCAATACACTGTTGATGCTATGCAAGAACGATGCTGAAT	717		
DB	621	GAAGATGCCACGAGTTCTTGAGGTACACTGTGATGCTATGCAAAAGTCTGCTGCTTCT	680		
QY	718	GGCAGCAATAAATTAGACAGACACACCAGGCCACCACTCTTGTTCAGATATTTGGA	777		
DB	681	G---GAAACAACTGGACAGCAAACTCAGGCAACCACTTTTGTACATCAGATTTTGA	737		
QY	778	GGATACCTAAGATCTAGAGTCAAAATGTTTAAATGCAAGGGCGTTCAGATACATTTT	837		
DB	738	GGATATCTGAAGTCCAGAGTCAAAATGCTGAAATGCAAGCAGTCTCTGATACCTTTC	797		
QY	838	CCATATCTTGATATAAATTTGAGATAAAGCTGCTCAGAGTGTCAACAGGCAATTTGAG	897		
DB	798	CCATATTTGGATATTTTCAATTCAGATAAAGCGGCTCAGACACTCTCTTAAGGCATTT	857		
QY	898	CAGTTTGTGAAGCCGGAACAGCTTTGATGGAGAAACTCGTCAAGTGGAGCAAGGTGAAA	957		
DB	858	CAATTTGTAAACCTGAGCAACTTGTATGGGACAAATGCTATAAATGCTCCAAATGTA	917		
QY	958	AAGATGGTTTCAGCTTCAAGAGGTTCACTATCCATAGATCTCTTAATGTTCTTACCTT	1017		
DB	918	AAATGGTTTCTGCTCTCAAGAGATTCACCGTTCATCGCAGTCTTAATGTTCTCAATC	977		

THIS PAGE BLANK (USPTO)

Result No.	Query %			ID	Description
	Score	Match	Length		
1	2412	100.0	2412	4	Aaf81717 Human pro
2	2371.6	98.3	4659	6	Abk86140 cDNA enco
3	2365	98.1	3660	4	Aah17633 Human cDN
4	2339.2	97.0	2347	4	Aah19322 Human ubi
5	2339.2	97.0	2347	7	Abx93880 Human cDN
6	2310	95.8	4483	6	Abz11508 Human pol
7	1675.6	69.5	1679	6	Abj39772 Human NS
8	727.2	30.1	16682	4	Aak79798 Human imm
9	727.2	30.1	16682	4	Aak70010 Human imm
10	727.2	30.1	21436	4	Aak79799 Human imm
11	727.2	30.1	21436	4	Aak70011 Human imm
12	517.4	21.5	1122	4	Aas41077 cDNA enco
13	493	20.4	577	4	Aas41613 cDNA enco
14	474.6	19.7	722	6	Abk30453 Human G-p
15	468	19.4	521	4	Aah05659 Human cDN
16	443	18.4	1593	9	Adc27036 Human deu
17	436	18.1	1698	9	Adc27022 cDNA enco
18	434.6	18.0	1763	7	ACA61698 Human deu
19	433.4	18.0	1593	9	Adc27020 Human deu
20	433	18.0	4746	4	Aah46504 Human deu
21	431.8	17.9	1593	9	Adc27014 Human deu
22	432.2	17.9	3051	9	Adc27060 Human deu
23	430.2	17.8	1593	9	Adc27028 Human deu

Db 1021 ACGTTTGGCAATTTTACCGTGGAAAATGCTAAGGATGTGAANTACCCCTGAGTATCT 1080
 QY 1086 TGATATTCGGCCCATATATGCTCAACCCACGAGAGCCAAATTTGTCTAAGTCTTGTATGC 1145
 Db 1081 TGATATTCGGCCCATATATGCTCAACCCACGAGAGCCAAATTTGTCTAAGTCTTGTATGC 1140
 QY 1146 AGTGTCTGGTCCACACTGGTTTAAATTTGCCATGCTGGCCATTAATCTTCTGTACATAAAAGC 1205
 Db 1141 AGTGTCTGGTCCACACTGGTTTAAATTTGCCATGCTGGCCATTAATCTTCTGTACATAAAAGC 1200
 QY 1206 TAGCAATGGCCCTCTGGTATCAAAATGAATGACTCCATTTGATATACAGAGTGAATTAAGATC 1265
 Db 1201 TAGCAATGGCCCTCTGGTATCAAAATGAATGACTCCATTTGATATACAGAGTGAATTAAGATC 1260
 QY 1266 GGTACTCAGCCAAACAGCCCTATGCTCTTTTATATACAGTCCCATGATGTAATAATGG 1325
 Db 1261 GGTACTCAGCCAAACAGCCCTATGCTCTTTTATATACAGTCCCATGATGTAATAATGG 1320
 QY 1326 AGGTGAATTAATCAATCCACCACTAGCCCGGCGAGTCTCTCCCGCCCGCTCATCAG 1385
 Db 1321 AGGTGAATTAATCAATCCACCACTAGCCCGGCGAGTCTCTCCCGCCCGCTCATCAG 1380
 QY 1386 TCAGCGGTTGTACCAACAAACAGGCTGCGCCAGGCTTTATCGACCAACAGCTTCCTC 1445
 Db 1381 TCAGCGGTTGTACCAACAAACAGGCTGCGCCAGGCTTTATCGACCAACAGCTTCCTC 1440
 QY 1446 TCATGATGAATGAATCACTCACTTAATGGAATGGACCTGGACCATTTGAAGACAGCCCAAG 1505
 Db 1441 TCATGATGAATGAATCACTCACTTAATGGAATGGACCTGGACCATTTGAAGACAGCCCAAG 1500
 QY 1506 CAGTTCCATGTCGAGTCTTAACGGGAATTCAGTGTCAACAGGGCTAGTCTGTTAATGC 1565
 Db 1501 CAGTTCCATGTCGAGTCTTAACGGGAATTCAGTGTCAACAGGGCTAGTCTGTTAATGC 1560
 QY 1566 TTCAGTCTCTGTCAAAACTGGTCAGTTAATAGTCTCTAGTGATCCAGAACATCTAA 1625
 Db 1561 TTCAGTCTCTGTCAAAACTGGTCAGTTAATAGTCTCTAGTGATCCAGAACATCTAA 1620
 QY 1626 GAAACAAAATTAACATCAGTATTCACAAAGTTGGCTTTCGCCAGTGTCAAGTCTCA 1685
 Db 1621 GAAACAAAATTAACATCAGTATTCACAAAGTTGGCTTTCGCCAGTGTCAAGTCTCA 1680
 QY 1686 ACCTAAGCTTCATAGTAAATCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCAT 1745
 Db 1681 ACCTAAGCTTCATAGTAAATCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCAT 1740
 QY 1746 TACCAATTCAGAGTACAGTCTACCTCGAAGCATCTAAGATGTCAAGTTCTAGTAAAGT 1805
 Db 1741 TACCAATTCAGAGTACAGTCTACCTCGAAGCATCTAAGATGTCAAGTTCTAGTAAAGT 1800
 QY 1806 AACAAAACCGATCCCGCCAGTGAATCTGTCCAGCCCGTGTGAATGCAATCCAA 1865
 Db 1801 AACAAAACCGATCCCGCCAGTGAATCTGTCCAGCCCGTGTGAATGCAATCCAA 1860
 QY 1866 GCTGAATCCAGCGTGTGCTGCTATGGCGCCGAGTCTCTGAGGACTCTGACGAGGA 1925
 Db 1861 GCTGAATCCAGCGTGTGCTGCTATGGCGCCGAGTCTCTGAGGACTCTGACGAGGA 1920
 QY 1926 GTCAAAAGGGCTGGGCAAGGAGATGGATTTGGTACGATTTGTAGCTCCCACTCTCCGG 1985
 Db 1921 GTCAAAAGGGCTGGGCAAGGAGATGGATTTGGTACGATTTGTAGCTCCCACTCTCCGG 1980
 QY 1986 CCAAGATCCGAAATGAGGAGGCACTCCCGACGAGCTTCGAACCATGACCCCTAAA 2045
 Db 1981 CCAAGATCCGAAATGAGGAGGCACTCCCGACGAGCTTCGAACCATGACCCCTAAA 2040
 QY 2046 CGGTCTTAATAGTGCAGACAGCAGTGAACCCGAAAGAAAACGGCCTAGCCCTGATGG 2105
 Db 2041 CGGTCTTAATAGTGCAGACAGCAGTGAACCCGAAAGAAAACGGCCTAGCCCTGATGG 2100
 QY 2106 TGCCAGCTGCCAAGCCAGCCTGCTGCTCAAGAAATCCCTTTGCTAAGGCAACGG 2165
 Db 2101 TGCCAGCTGCCAAGCCAGCCTGCTGCTCAAGAAATCCCTTTGCTAAGGCAACGG 2160

QY 2166 TCTTCTCGAAGTTGATGCTCTCTCTTTGCTGTCTCTCCAGAAACAAAAATCTTAGA 2225
 Db 2161 TCTTCTCGAAGTTGATGCTCTCTCTTTGCTGTCTCTCCAGAAACAAAAATCTTAGA 2220
 QY 2226 GACCTTCAGCTTGAACAACAACTGAAAGCTCGAGGATGAATGAGTGCACCTGGAGC 2285
 Db 2221 GACCTTCAGCTTGAACAACAACTGAAAGCTCGAGGATGAATGAGTGCACCTGGAGC 2280
 QY 2286 AGAGAGGGCCCTCCGAGGACCGCGACGCGAGCTTCAGCTTGGCAGCCCGCCGCCGA 2345
 Db 2281 AGAGAGGGCCCTCCGAGGACCGCGACGCGAGCTTCAGCTTGGCAGCCCGCCGCCGA 2340
 QY 2346 ATCCCTGGAGGACGAGATGCGGCCGCAAGCTT 2378
 Db 2341 ATCCCTGGAGGACGAGATGCGGCCGCGCCCT 2373
 RESULT 4
 AAH19322
 ID AAH19322 standard; cDNA; 2347 BP.
 XX AAH19322;
 XX 20-JUL-2001 (first entry)
 XX Human ubiquitin protease 23431 coding sequence.
 DE Human; ubiquitin protease; deubiquitinating enzyme; tuberculostatic;
 KW antiasthmatic; antiinflammatory; antidiarrhoeic; hepatotropic;
 KW gynaecological; cytostatic; antimicrobial; neuroprotective; anti-HIV;
 KW immunosuppressive; cardiac; antianaemic; nephrotropic; antibacterial;
 KW anti-thyroid; gastroin-estinal; ss.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FH 61..2346
 CDS /*tag= a
 FT /partial
 FT /product= "Human ubiquitin protease"
 FT /note= "No stop codon given"
 XX WO200123589-A2.
 PN 05-APR-2001.
 XX 29-SEP-2000; 2000WO-US026962.
 PF 29-SEP-1999; 99US-00407356.
 PR (MILL-) MILLENNIUM PHARM INC.
 XX Kapeller-Libermann R;
 PI WPI; 2001-374253/39.
 XX P-PSDB; AAB82177.
 XX New human ubiquitin protease, a member of the mammalian deubiquitinating
 PT enzymes is useful for diagnosis and treatment of e.g. tuberculosis and
 PT Alzheimer's disease.
 XX
 PS Claim 1; Fig 1; 111pp; English.
 XX
 CC The present sequence is the coding sequence for human ubiquitin protease
 CC 23431. The ubiquitin protease coding sequence and protein are useful as
 CC reagents or targets in assays for treatment and diagnosis of ubiquitin-
 CC mediated or related disorders, especially disorders mediated by
 CC deubiquitinating enzymes. The protein and coding sequence are also useful
 CC for treating disorders involving the following: the spleen e.g.
 CC tuberculosis and typhoid fever, the lung such as bronchial asthma, the
 CC colon e.g. as diarrhoea and dysentery, the liver e.g. jaundice and
 CC cholestasis, the uterus and endometrium e.g. endometritis, the brain e.g.

Db	1864	ACGGTCTGGTCCCTATGGCCGAGTCTCTGAGACTCTGACGAGGATCAAGGGG	1923
QY	1936	CTGGGCAAGGAGAAATGGGATTTGATGCTGAGCTCCACTCTCCCGCCAAAGATGCC	1995
Db	1924	CTGGGCAAGGAGAAATGGGATTTGATGCTGAGCTCCACTCTCCCGCCAAAGATGCC	1983
QY	1996	GAAGATGAGAGGCCACTCCGCGACGAGCTTCAAGAACCCATGACCTTAACGGTGTAAAT	2055
Db	1984	GAAGATGAGAGGCCACTCCGCGACGAGCTTCAAGAACCCATGACCTTAACGGTGTAAAT	2043
QY	2056	AGTGCACAGCAGCAGTACCGGAAAGAAACGGCTAGCGCTGATGTGTCAGCTGC	2115
Db	2044	AGTGCACAGCAGCAGTACCGGAAAGAAACGGCTAGCGCTGATGTGTCAGCTGC	2103
QY	2116	CAAGGCGAGCGCTCCCTGCACTCAGAAATCCCTTTGCTAAGGCAACGGTCTTCTGGA	2175
Db	2104	CAAGGCGAGCGCTCCCTGCACTCAGAAATCCCTTTGCTAAGGCAACGGTCTTCTGGA	2163
QY	2176	AAGTTGATGCTGCTCTTGTCTCTCCAGAGCAAAATCTTAGAGACTTCCAGG	2235
Db	2164	AAGTTGATGCTGCTCTTGTCTCTCCAGAGCAAAATCTTAGAGACTTCCAGG	2223
QY	2236	CTTAGCAACAACTGAAGGCTCGACGATGAATGAGTGCACCTGGAGCAGAGAGGGC	2295
Db	2224	CTTAGCAACAACTGAAGGCTCGACGATGAATGAGTGCACCTGGAGCAGAGAGGGC	2283
QY	2296	CCTCCGAGGACCGCGAGCGCTCAGCTGCGAGCCCGCCCGGCGAATCCTCTGAG	2355
Db	2284	CCTCCGAGGACCGCGAGCGCTCAGCTGCGAGCCCGCCCGGCGAATCCTCTGAG	2343
QY	2356	GAGC 2359	
Db	2344	GAGC 2347	
RESULT 5			
ID	ABX93880	standard; cDNA; 2347 BP.	
XX	AC	ABX93880;	
XX	DT	10-JUN-2003 (first entry)	
XX	DE	Human cDNA encoding ubiquitin protease.	
XX	KW	Human; ss; gene; ubiquitin protease; spleen disorder; Hodgkin's disease;	
XX	KW	lung disorder; adult respiratory distress syndrome; colon disorder;	
XX	KW	inflammatory bowel disease; liver disorder; jaundice; uterine disorder;	
XX	KW	endometriosis; brain disorder; Alzheimer's disease; kidney disorder;	
XX	KW	acquired immunodeficiency syndrome; AIDS; skin disorder; urticaria;	
XX	KW	heart disorder; ischaemic heart disease; blood vessel disorder;	
XX	KW	atherosclerosis; red blood cell disorder; anaemia; thymus disorder;	
XX	KW	DiGeorge syndrome; B-cell disorder; leukaemia; kidney disorder;	
XX	KW	polycystic kidney disease; glomerulonephritis; breast disorder; mastitis;	
XX	KW	testicular disorder; sexually transmitted disease; thyroid disorder;	
XX	KW	hypothyroidism; pancreatic disorder; pancreatitis; intestinal disorder;	
XX	KW	whipple disease; tumour; cancer.	
OS	Homo sapiens.		
XX	Key	Location/Qualifiers	
XX	FT	61..2347	
XX	FT	/tag= a	
XX	FT	/product= "Ubiquitin protease"	
XX	FT	/partial	
XX	FT	/note= "No stop codon shown"	
XX	US2003037350-A1.		
XX	PN		
XX	XX		
XX	PD	20-FEB-2003.	
XX	XX		
XX	PF	05-JUN-2002; 2002US-00163547.	

PR	29-SEP-1999;	99US-00407356.	
PR	05-NOV-1999;	99US-00435311.	
PR	01-FEB-2000;	2000US-00496005.	
PR	11-FEB-2000;	2000US-0182009P.	
PR	14-FEB-2000;	2000US-0182408P.	
PR	28-FEB-2000;	2000US-0185503P.	
PR	02-NOV-2000;	2000US-00704918.	
PR	12-FEB-2001;	2001US-00781598.	
PR	14-FEB-2001;	2001US-00782952.	
PR	28-FEB-2001;	2001US-00796100.	
XX	(MILL-) MILLENNIUM PHARM INC.		
PA	Glucksmann MA, Kapeller-Libermann R, Meyers RE, Rudolph-Owen LA;		
XX	WPI; 2003-342047/32.		
XX	P-PSDB; ABU08951.		
XX	New nucleic acid or polypeptide, useful for preparing a composition for		
XX	diagnosing or treating diseases e.g., tumor.		
XX	Claim 2; Fig 1; 281pp; English.		
XX	The invention relates to a new isolated nucleic acid molecule encoding		
XX	one of 6 polypeptides (ubiquitin protease, lipase, dynamin, short chain		
XX	dehydrogenase, ADAM-TS (a disintegrin and metalloprotease domain protein		
XX	with thrombospondin (TS) domains) and gamma butyrobetaine-hydroxylase		
XX	(gamma BH). Also included are the polypeptide, host cells containing the		
XX	nucleic acid, an antibody that selectively binds to the polypeptide, a		
XX	method for producing the polypeptide, a method for detecting the presence		
XX	of the polypeptide or the nucleic acid in a sample, a method for		
XX	identifying a compound that binds to the polypeptide. The nucleic acid		
XX	modulating the activity of the polypeptide and a method for diagnosing or		
XX	or polypeptide is useful for preparing a composition for diagnosing or		
XX	treating diseases e.g. spleen disorders (e.g. splenomegaly and Hodgkin's		
XX	disease), lung disorders (e.g. adult respiratory distress syndrome,		
XX	pulmonary oedema, chronic bronchitis and emphysema), colon disorders		
XX	(e.g. stenosis, colitis, inflammatory bowel disease and Crohn's disease),		
XX	liver disorders (e.g. jaundice, cirrhosis, hepatitis and alcoholic liver		
XX	disease), uterine and endometrial disorders (e.g. endometriosis and		
XX	menopausal changes), brain disorders (e.g. encephalitis, Alzheimer's		
XX	disease, Parkinson's disease, ataxia and multiple sclerosis), T-cell		
XX	disorders (e.g. acquired immunodeficiency syndrome, AIDS), skin disorders		
XX	(e.g. urticaria, dermatitis and lupus erythematosus), heart disorders		
XX	(e.g. ischaemic heart disease, myocardial infarction and cardiomyopathy),		
XX	blood vessel disorders (e.g. atherosclerosis, thrombophlebitis and		
XX	Raynaud disease), red blood cell disorders (e.g. anaemia), thymus		
XX	disorders (e.g. DiGeorge syndrome), B-cell disorders (e.g. leukaemia),		
XX	kidney disorders (e.g. polycystic kidney disease and glomerulonephritis),		
XX	breast disorders (e.g. mastitis), testicular disorders (e.g. sexually		
XX	transmitted diseases and cryptorchidism), thyroid disorders (e.g.		
XX	hypothyroidism), pancreatic disorders (e.g. pancreatitis), and intestinal		
XX	disorders (e.g. whipple disease), as well as tumours and cancers of the		
XX	above listed organs/cells. Many more diseases and disorders are listed in		
XX	the specification. The present sequence encodes human ubiquitin protease		
XX	Sequence 2347 BP; 647 A; 606 C; 528 G; 566 T; 0 U; 0 Other;		
SQ	Query Match	57.0%; Score 2339.2; DB 7; Length 2347;	
SQ	Best Local Similarity	59.9%; Pred. No. 0;	
SQ	Matches 2341; Conservative	0; Mismatches 3; Indels 0; Gaps 0;	
QY	16	GGCGCGGCGCGCGCGGCGGATGCGAGCGCGCGCGCGCGGTCAGAGTTGAACAATG	75
Db	4	CGTCCGCGGCGCGCGCGGCGGATGCGAGCGCGCGCGCGGTCAGAGTTGAACAATG	63
QY	76	ACCATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCTATCAGATCAGCTGCG	135
Db	64	ACCATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCTATCAGATCAGCTGCG	123
QY	136	AGCTCGAGGCGAGTCTCAGCTGGAGCATGGATGAGGTTCTGCCAGCTGGGTGCTGTG	195

Db 124 AGCTCGAGCGAGTCTCACCTGGAGACATGGATGCAGGTTCTGCCAGCTGGGGTGCTGTG 183
QY 196 TCTTCAATCGAATCATCTGTCAAAATCACACATTTCTTTAGGACAGTACCTGTGTCTGTA 255
Db 184 TCTTCAATGAAATGATGTGCAAAATCACACATTTCTTTAGGACAGTACCTGTGTCTGTA 243
QY 256 GTTATTTCGAGTTTCATCTGTACCTGTATAATCAAAACCATCACCAAAAGGATCAAGCC 315
Db 244 GTTATTTCGAGTTTCATCTGTACCTGTATAATCAAAACCATCACCAAAAGGATCAAGCC 303
QY 316 CTAGGTGATGGCANCTCTCTCCACAGAAAGTTCTTTTCCCATCTGAGAGATTTTGTCTT 375
Db 304 CTAGGTGATGGCATCTCTCTCCACAGAAAGTTCTTTTCCCATCTGAGAGATTTTGTCTT 363
QY 376 AAGTGGCAACAACTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGGCAATACCTGTTTT 435
Db 364 AAGTGGCAACAACTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGGCAATACCTGTTTT 423
QY 436 GCCAATGACGACTGAGTGTGTTTAACTTACACACACACCTCTTGGCCAAATTACATGCTATCA 495
Db 424 GCCAATGACGACTGAGTGTGTTTAACTTACACACACACCTCTTGGCCAAATTACATGCTATCA 483
QY 496 CATGAACACTCOAAACATGTGATGAGAGGCTTTTGTATGATGTGTACAAATGCAAGCA 555
Db 484 CATGAACACTCOAAACATGTGATGAGAGGCTTTTGTATGATGTGTACAAATGCAAGCA 543
QY 556 CATATTACCCAGGCACTCAGTAATCCTGGGGAGGTTATTAACCAATGTTGTGATCAAT 615
Db 544 CATATTACCCAGGCACTCAGTAATCCTGGGGAGGTTATTAACCAATGTTGTGATCAAT 603
QY 616 GAGATGCGGCGTATAGTAGGCACTCTCGTTTGGAAACCAAGAGATGCCCATGAATTC 675
Db 604 GAGATGCGGCGTATAGTAGGCACTCTCGTTTGGAAACCAAGAGATGCCCATGAATTC 663
QY 676 CTTCATATCAGTGTGATGCTATGAGAAAGCATGCTTGAATGGCAGCAATAAATTAGAC 735
Db 664 CTTCATATCAGTGTGATGCTATGAGAAAGCATGCTTGAATGGCAGCAATAAATTAGAC 723
QY 736 AGACACACCCAGGCCACCACTCTTGTGTGATATTTGGAGGATACCTAAGATCTAGA 795
Db 724 AGACACACCCAGGCCACCACTCTTGTGTGATATTTGGAGGATACCTAAGATCTAGA 783
QY 796 GTCAAAATGTTTAAATTTGCAAGGGCGTTTCAGATACCTTTTGATCCATATCTTGATATAACA 855
Db 784 GTCAAAATGTTTAAATTTGCAAGGGCGTTTCAGATACCTTTTGATCCATATCTTGATATAACA 843
QY 856 TTGAGATAAAGGCTGTCTAGAGTGTCAACAAGGCATTTGGAGCAGTTTGTGAAGCCGGAA 915
Db 844 TTGAGATAAAGGCTGTCTAGAGTGTCAACAAGGCATTTGGAGCAGTTTGTGAAGCCGGAA 903
QY 916 CAGCTTGATGAGAAAACTCGTACAAAGTGCAGCAGTGTAAAGATGGTTCCAGCTTCA 975
Db 904 CAGCTTGATGAGAAAACTCGTACAAAGTGCAGCAGTGTAAAGATGGTTCCAGCTTCA 963
QY 976 AAGAGGTTCACTATCCATAGATCCCTTAATGTTCTTACACTTCTCTGAAACGTTTTCGA 1035
Db 964 AAGAGGTTCACTATCCATAGATCCCTTAATGTTCTTACACTTCTCTGAAACGTTTTCGA 1023
QY 1036 AATTTTACCGTGGAAAAATTTGCTAAGGATGTGAAATACCTGTAGTATCTTGTATTCGG 1095
Db 1024 AATTTTACCGTGGAAAAATTTGCTAAGGATGTGAAATACCTGTAGTATCTTGTATTCGG 1083
QY 1096 CCATATATGCTCAACCAACGGGAGGCCAATTTGTCTACGTTCTGTATGCAAGTGTGCTC 1155
Db 1084 CCATATATGCTCAACCAACGGGAGGCCAATTTGTCTACGTTCTGTATGCAAGTGTGCTC 1143
QY 1156 CACACTGGTTTAAATTTGCCATGCTGGCCATTTCTCTGCTACATAAAGCTAGCAATGGC 1215
Db 1144 CACACTGGTTTAAATTTGCCATGCTGGCCATTTCTCTGCTACATAAAGCTAGCAATGGC 1203
QY 1216 CTCTGGTATCAAAATGAATGACTCCATTGTATCTACAGTGTATTTAGATCGGTACTCAGC 1275
Db 1204 CTCTGGTATCAAAATGAATGACTCCATTGTATCTTACAGTGTATTTAGATCGGTACTCAGC 1263

QY 1276 CAAACAAGCCTATGTGCTCTTTTATATCAGTCCCATGATGTGAAAAATGAGGTGAACCT 1335
Db 1264 CAAACAAGCCTATGTGCTCTTTTATATCAGTCCCATGATGTGAAAAATGAGGTGAACCT 1323
QY 1336 ACTCATCCCAACCATAGCCCCGGGCCAGTCTCTCCCGGCCCTCATCTAGTCAGCGGGTT 1395
Db 1324 ACTCATCCCAACCATAGCCCCGGGCCAGTCTCTCTCCCGGCCCTCATCTAGTCAGCGGGTT 1383
QY 1396 GTCAACCAACAAACAGGCTCGCCAGGCTTTTATCGGACCAACAGCTTCCCTCTCACATGATA 1455
Db 1384 GTCAACCAACAAACAGGCTCGCCAGGCTTTTATCGGACCAACAGCTTCCCTCTCACATGATA 1443
QY 1456 AAGAATCCACCTCATCTTAAATGGGACTGCACCAATTGAAAGACACGCAAGCAGTTCATG 1515
Db 1444 AAGAATCCACCTCACTTAAATGGGACTGCACCAATTGAAAGACACGCAAGCAGTTCATG 1503
QY 1516 TCGAGTCCCTAAACGGGAATTCAGTGTCAACAGGGCTAGTCTGTTAATGCTTCAGCTTCT 1575
Db 1504 TCGAGTCCCTAAACGGGAATTCAGTGTCAACAGGGCTAGTCTGTTAATGCTTCAGCTTCT 1563
QY 1576 GTCCAAAACTGTCTCAGTTTAAATAGGTCTCTCAGTGTATCCAGAGACATCCTAAGAAACAAAA 1635
Db 1564 GTCCAAAACTGTCTCAGTTTAAATAGGTCTCTCAGTGTATCCAGAGACATCCTAAGAAACAAAA 1623
QY 1636 ATTAACAATCAGTATTCACAAACAGTTGCTGTTGCGCAGTGTCACTCAACTAACCTT 1695
Db 1624 ATTAACAATCAGTATTCACAAACAGTTGCTGTTGCGCAGTGTCACTCAACTAACCTT 1683
QY 1696 CATAGTAAATCTTTGCGAGAACCTTACCAAGCCGTTCCCTCTTACGATTAACCAATTC 1755
Db 1684 CATAGTAAATCTTTGCGAGAACCTTACCAAGCCGTTCCCTCTTACCAATTACCAATTC 1743
QY 1756 GCAGTACAGTCTACCTCGAAACGATCTACGATGTCACTTCTAGTAAAGTAAACAAACCG 1815
Db 1744 GCAGTACAGTCTACCTCGAAACGATCTACGATGTCACTTCTAGTAAAGTAAACAAACCG 1803
QY 1816 ATCCCCCGCAGTGAATCTCTGCTCCAGGCCCTGATGAATGGCAAAATCAAAGCTGAACCTC 1875
Db 1804 ATCCCCCGCAGTGAATCTCTGCTCCAGGCCCTGATGAATGGCAAAATCAAAGCTGAACCTC 1863
QY 1876 AGCGTCTGTTGCCCTATGGGCCGAGTCTCTTGAGGACTCTGACGAGAGTCAAAAGGG 1935
Db 1864 AGCGTCTGTTGCCCTATGGGCCGAGTCTCTTGAGGACTCTGACGAGAGTCAAAAGGG 1923
QY 1936 CTGGCAAGAGAGAAATGGGATTTGATGATTTGTAGTCTCCACTCTCCCGGCCCAAGATGCC 1995
Db 1924 CTGGCAAGAGAGAAATGGGATTTGATGATTTGTAGTCTCCACTCTCCCGGCCCAAGATGCC 1983
QY 1996 GAAAGATGAGGAGGCACTCCGCAAGCTTCAAGAACCCATGACCTCAAAACGCTCTAAT 2055
Db 1984 GAAAGATGAGGAGGCACTCCGCAAGCTTCAAGAACCCATGACCTCAAAACGCTCTAAT 2043
QY 2056 AGTCAGACGCGACAGTGCACCCGAAAGAAAACGGCTAGCGCTGTATGGTGCAGTGC 2115
Db 2044 AGTCAGACGCGACAGTGCACCCGAAAGAAAACGGCTAGCGCTGTATGGTGCAGTGC 2103
QY 2116 CAAGGCCAGCTGCTGCTCACTCAGAAAATCCCTTTGCTAAGGCAAAACGGCTTCTCTGGA 2175
Db 2104 CAAGGCCAGCTGCTGCTCACTCAGAAAATCCCTTTGCTAAGGCAAAACGGCTTCTCTGGA 2163
QY 2176 AAGTTGATGCTGCTCTTGTCTCTCTCCAGAGACAAAATCTTAGAGACCTTCAGG 2235
Db 2164 AAGTTGATGCTGCTCTTGTCTCTCTCCAGAGACAAAATCTTAGAGACCTTCAGG 2223
QY 2236 CTTAGCAACAACTGAAGAGCTGACGAGTGAATGATGTGACCTGGAGCAGAGGGGC 2295
Db 2224 CTTAGCAACAACTGAAGAGCTGACGAGTGAATGATGTGACCTGGAGCAGAGGGGC 2283
QY 2296 CCTCCGAGACCGCAGCGAGCTCAGCTGGAGCCCCCGCCGGAATCCCTCGAG 2355
Db 2284 CCTCCGAGACCGCAGCGAGCTCAGCTGGAGCCCCCGCCGGAATCCCTCGAG 2343

CC	diseases, dystonia, multiple sclerosis, rheumatoid arthritis, psoriasis,
CC	cataracts, restenosis, atherosclerosis, inflammation, skin disorders,
CC	glaucoma, obesity, muscular dystrophy, AIDS, infertility, cardiovascular
CC	disease, coagulation disease, ischemia, hypertension, asthma, immune
CC	disease, epilepsy, angina, neurodegeneration, diabetes, anxiety,
CC	depression, schizophrenia, viral disease, gastric ulcers, stroke,
CC	Alzheimer's disease and as a contraceptive
XX	
XX	Sequence 1679 BP; 461 A; 443 C; 378 G; 393 T; 0 U; 4 Other;
QY	Query Match 69.5%; Score 1675.6; DB 6; Length 1679;
DB	Best Local Similarity 99.8%; Pred. No. 0;
QY	Matches 1675; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
DB	
QY	694 GCTATGACAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCACC 753
DB	1 GCTATGACAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCACC 60
QY	754 ACTCTTGTGTGTCAGATATTGAGAGATACCTAAGATCTAGAGTCAAAATGTTAAATTGC 813
DB	61 ACTCTTGTGTGTCAGATATTGAGAGATACCTAAGATCTAGAGTCAAAATGTTAAATTGC 120
QY	814 AAGGCGCTTTCAGATATTGATCCATATCTTCATATAAATTGGAGATAAAGGCTGCT 873
DB	121 AAGGCGCTTTCAGATATTGATCCATATCTTCATATAAATTGGAGATAAAGGCTGCT 180
QY	874 CAGAGTGTCAACAAGGCATTGGAGCAGTTTGTGAAGCGGAAACAGCTTGATGGAGAAAC 933
DB	181 CAGAGTGTCAACAAGGCATTGGAGCAGTTTGTGAAGCGGAAACAGCTTGATGGAGAAAC 240
QY	934 TCGTACAAAGTCAGCAAGTGTAAAGAGTGTTCAGCTTCAAGAGGTTCACTATCCAT 993
DB	241 TCGTACAAAGTCAGCAAGTGTAAAGAGTGTTCAGCTTCAAGAGGTTCACTATCCAT 300
QY	994 AGATCCCTAATGTTCTTACACTTCTCTGAAACGTTTGTGAAATTTTACCGGTGGAATA 1053
DB	301 AGATCCCTAATGTTCTTACACTTCTCTGAAACGTTTGTGAAATTTTACCGGTGGAATA 360
QY	1054 ATTGTAAGGATGTGAATACCTGAGTATCTTGATATTCGGCCCATATGTTCTCAACC 1113
DB	361 ATTGTAAGGATGTGAATACCTGAGTATCTTGATATTCGGCCCATATGTTCTCAACC 420
QY	1114 AACGGAGAGCAATTGTCATGCTTGTGTATGAGCTGCTGGTCCACACTGGTTTAAATTGC 1173
DB	421 AACGGAGAGCAATTGTCATGCTTGTGTATGAGCTGCTGGTCCACACTGGTTTAAATTGC 480
QY	1174 CATGCTGGCCATTACTTCTGCTACATAAAGCTAGCAATGGCTCTGGTATCAAAATGAAT 1233
DB	481 CATGCTGGCCATTACTTCTGCTACATAAAGCTAGCAATGGCTCTGGTATCAAAATGAAT 540
QY	1234 GACTCCATTGATCTACCAAGTATATTAGATCGGTACTCAGCCAAAGCCCTATGTGCTC 1293
DB	541 GACTCCATTGATCTACCAAGTATATTAGATCGGTACTCAGCCAAAGCCCTATGTGCTC 600
QY	1294 TTTTATATCAGTCCCATGATGTGAAATATGAGTGAATCTACTCATCCACCCATAGC 1353
DB	601 TTTTATATCAGTCCCATGATGTGAAATATGAGTGAATCTACTCATCCACCCATAGC 660
QY	1354 CCGGCGCAGTCTCTCCCGCCCGCTCATCAGTCAGCGGGTGTGTCAACCAACAAAGGCT 1413
DB	661 CCGGCGCAGTCTCTCTCCCGCCCGCTCATCAGTCAGCGGGTGTGTCAACCAACAAAGGCT 720
QY	1414 GCGCCAGGCTTTATCGGACACAGCTTCCCTCTCATATGATAAAGATCCACCTCACTTA 1473
DB	721 GCGCCAGGCTTTATCGGACACAGCTTCCCTCTCATATGATAAAGATCCACCTCACTTA 780
QY	1474 AATGGAGCTGGACCATGTGAAGACACCGCAAGCAGTTCCATGTCGAGTCCCTAACCGGAAT 1533
DB	781 AATGGAGCTGGACCATGTGAAGACACCGCAAGCAGTTCCATGTCGAGTCCCTAACCGGAAT 840
QY	1534 TCCAGTGTCAACAGGGCTAGTCTCTGTTAAATGCTTTCAGCTTCTGTGCCAAAATCGGT 1593
DB	841 TCCAGTGTCAACAGGGCTAGTCTCTGTTAAATGCTTTCAGCTTCTGTGCCAAAATCGGT 900

QY	1594 AATAGGTCTCAGTGTATCCAGAAACATCTTAAGAAAACAAAAATTACAATCAGTATTAC 1653
DB	901 AATAGGTCTCAGTGTATCCAGAAACATCTTAAGAAAACAAAAATTACAATCAGTATTAC 960
QY	1654 AACAAAGTTCGCTGTTCCGAGTGTGATCTCAACCTTAACCTTCATAGTAAATTTTGGAG 1713
DB	961 AACAAAGTTCGCTGTTCCGAGTGTGATCTCAACCTTAACCTTCATAGTAAATTTTGGAG 1020
QY	1714 AACCTTACCAAGCCGTTCCCTCTCTTACCATTAACCAATCTTCAGTACAGTCTACCTCG 1773
DB	1021 AACCTTACCAAGCCGTTCCCTCTCTTACCATTAACCAATCTTCAGTACAGTCTACCTCG 1080
QY	1774 AAGCATCTACGAATGTCAGTCTTAGTAAAGTAACAAAACCGATCCCGCAGTGAATCC 1833
DB	1081 AAGCATCTACGAATGTCAGTCTTAGTAAAGTAACAAAACCGATCCCGCAGTGAATCC 1140
QY	1834 TGTCTCCAGCCGCTGATGAATGGCAATCCAGCTGAATCCAGCTGAGTGTGGTCCCTAT 1893
DB	1141 TGTCTCCAGCCGCTGATGAATGGCAATCCAGCTGAATCCAGCTGAGTGTGGTCCCTAT 1200
QY	1894 GCGCCGAGTCTCTGAGGACTCTGACGAGAGTCAAGGGGCTGGCAAGGAGAAATGGG 1953
DB	1201 GCGCCGAGTCTCTGAGGACTCTGACGAGAGTCAAGGGGCTGGCAAGGAGAAATGGG 1250
QY	1954 ATTGTTACGATTTGAGTCTCCACTCTCCCGGCCAAGATGCCGAAGATGAGGAGGCCACT 2013
DB	1261 ATTGTTACGATTTGAGTCTCCACTCTCCCGGCCAAGATGCCGAAGATGAGGAGGCCACT 1320
QY	2014 CCGCAGAGCTTCAAGAACCCATGACCTTAACCGTCTTAATAGTGACAGCAGCAGACAGT 2073
DB	1321 CCGCAGAGCTTCAAGAACCCATGACCTTAACCGTCTTAATAGTGACAGCAGCAGACAGT 1380
QY	2074 GACCGGAGAAAGAGCGCTAGCGCTGATGCTGACCTGACCTGCAAGGCGCAGCTGCCCTG 2133
DB	1381 GACCGGAGAAAGAGCGCTAGCGCTGATGCTGACCTGACCTGCAAGGCGCAGCTGCCCTG 1440
QY	2134 CACTCAGAAAAATCCCTTTGCTTAAGGCAAAACCGTCTTCTGGAAGTTGATGCTGCTCT 2193
DB	1441 CACTCAGAAAAATCCCTTTGCTTAAGGCAAAACCGTCTTCTGGAAGTTGATGCTGCTCT 1500
QY	2194 TTGCTGTCTCTCCAGAAAGACAAATCTTAGAGCTTTAGAGCTTAGCAACAACTGAAA 2253
DB	1501 TTGCTGTCTCTCCAGAAAGACAAATCTTAGAGCTTTAGAGCTTAGCAACAACTGAAA 1560
QY	2254 GGCTCGAGGATGAATCAGTGCACCTGAGCAGAGAGGGGCTCCCGAGGACCGCGAC 2313
DB	1561 GGCTCGAGGATGAATCAGTGCACCTGAGCAGAGAGGGGCTCCCGAGGACCGCGAC 1620
QY	2314 GCGGAGCCTCAGCTTGCGAGCCCGCCCGCGAATCCCTGGAGGAGCAGATCGGCGCG 2372
DB	1621 GCGGAGCCTCAGCTTGCGAGCCCGCCCGCGAATCCCTGGAGGAGCAGATCGGCGCG 1679

RESULT 8

AAK79798

ID AAK79798 standard; DNA; 16682 BP.

XX AAK79798;

AC AAK79798;

XX 07-NOV-2001 (first entry)

DT Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34610.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX cytosolic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

OS WO200157182-A2.

PN XX

XX 09-AUG-2001.

PD XX

PF 17-JAN-2001; 2001WO-US001354.
 XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226688P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 21-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.

PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-02559678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,


```
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0233066P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239355P.
PR 13-OCT-2000; 2000US-0239377P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 08-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and metastasis.
Disclosure; SEQ ID NO 24822; 307lpp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
activity, and can be used in gene therapy and vaccine production. (I)
proteins and polynucleotides may be used in the prevention, diagnosis and
treatment of diseases associated with inappropriate (I) expression. For
example, they may be used to treat disorders associated with decreased
expression by rectifying mutations or deletions in a patient's genome
that affect the activity of (I) by expressing inactive proteins or to
supplement the patients own production of (I). Additionally, (I)
polynucleotides may be used to produce the secreted (I), by inserting the
nucleic acids into a host cell and culturing the cell to express the
protein. (I) proteins and polynucleotides may be used to prevent the
diagnosis and treat immune/haematopoietic-related diseases, especially
cancers and cancer metastases of haematopoietic-derived cells. AAK64703
to AAK87694 represent human immune/haematopoietic antigen genomic
sequences from the present invention. AAK54942 to AAK54950 and AAK82169
represent sequences used in the exemplification of the present invention

SQ Sequence 16682 BP; 4217 A; 3754 C; 4167 G; 4538 T; 0 U; 6 Other;

Query Match 30.1%; Score 727.2; DB 4; Length 16682;
Best Local Similarity 98.3%; Pred No. 9e-204;
Matches 735; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1435 CAGCTTCCCTTCACATGATAAGAAATCCACTCATTAAATGGGACTGGACATTGAAA 1494
| | | | |
Db 9429 CAGGTGCGCTCTGCTTACTTCCAGAATCCACTCATTAAATGGGACTGGACATTGAAA 9488
| | | | |

QY 1495 GACACGCCAAGCAGTTCATGTCGAGTCTTAACGGGAATTCAGTGTCAACAGGGCTAGT 1554
| | | | |
Db 9489 GACACGCCAAGCAGTTCATGTCGAGTCTTAACGGGAATTCAGTGTCAACAGGGCTAGT 9548
| | | | |

QY 1555 CCTGTTAATGCTTCAGCTTCTGTCCAAAACCTGCTCAGTTAATAGTCTCTCAGTGTATCCCA 1614
| | | | |
Db 9549 CCTGTTAATGCTTCAGCTTCTGTCCAAAACCTGCTCAGTTAATAGTCTCTCAGTGTATCCCA 9608
| | | | |

QY 1615 GAACATCCTAAGAAACAAAAAATTACAAATCAGTATTACAAACAAGTTCCCTGTTGCGCAG 1674
| | | | |
Db 9609 GAACATCCTAAGAAACAAAAAATTACAAATCAGTATTACAAACAAGTTCCCTGTTGCGCAG 9668
| | | | |

QY 1675 TGTGAGTCTCACTAACCTTTCATAGTAATCTTTGAGAGAACCTTACCAAGCCGGTTCCC 1734
| | | | |
```

RESULT 10

PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 08-NOV-2000; 2000US-0249207P.
 PR 08-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0251988P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 06-DEC-2000; 2000US-0256719P.
 PR 08-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 (HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI, 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis.

Disclosure; SEQ ID NO 34611; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patients own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/haematopoietic-related diseases, especially cancers and cancer metastases of haematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/haematopoietic antigen genomic

CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 21436 BP; 5267 A; 4950 C; 5418 G; 5801 T; 0 U; 0 Other;
 Query Match 30.1%; Score 727.2; DB 4; Length 21436;
 Best Local Similarity 98.3%; Pred. No. 1e-203;
 Matches 735; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
 QY 1435 CAGCTTCCCTCTCACATGATAAAGATCCACCTCCTTAATCGGAGTGGACCATTTGAAA 1494
 |||||
 Db 9432 CAGGTGGCTCTGCTTACTTCCAGAAATCCACCTACTTAATGGAGTGGACCATTTGAAA 9491
 |||||
 QY 1495 GACACGCCAAGAGTTCCATGTGAGTCTCTAAACGGGAAATCCAGTGTCAACAGGGGTAGT 1554
 |||||
 Db 9492 GACACGCCAAGAGTTCCATGTGAGTCTCTAAACGGGAAATCCAGTGTCAACAGGGGTAGT 9551
 |||||
 QY 1555 CTTGTTAATGCTTCAGCTTCTGTCGAAAATGCTGAGTTAATAGGTCCTCAGTGATCCCA 1614
 |||||
 Db 9552 CTTGTTAATGCTTTCAGCTTCTGTCGAAAATGCTGAGTTAATAGGTCCTCAGTGATCCCA 9611
 |||||
 QY 1615 GAACATCTTAAGAAAACAAAAAATTACAATCAGTATTCAACAAGTTGCTGTTCGCCAG 1674
 |||||
 Db 9612 GAACATCTTAAGAAAACAAAAAATTACAATCAGTATTCAACAAGTTGCTGTTCGCCAG 9671
 |||||
 QY 1675 TGTGAGTCTCAACCTTAACCTTCTTAGTAATTTCTTTGGAGAACCTTACCAAGCCGTTCC 1734
 |||||
 Db 9672 TGTGAGTCTCAACCTTAACCTTCTTAGTAATTTCTTTGGAGAACCTTACCAAGCCGTTCC 9731
 |||||
 QY 1735 TCTTCTACCAATTACCAATTTCTGAGTACAGTCTACTCTGAAACGATCTACGATGTGAGT 1794
 |||||
 Db 9732 TCTTCTACCAATTACCAATTTCTGAGTACAGTCTACTCTGAAACGATCTACGATGTGAGT 9791
 |||||
 QY 1795 TCTAGTAAAGTAACAAAACCGATCCCGCAGTGAATCTGCTCCAGCCCGGTGATGAAT 1854
 |||||
 Db 9792 TCTAGTAAAGTAACAAAACCGATCCCGCAGTGAATCTGCTCCAGCCCGGTGATGAAT 9851
 |||||
 QY 1855 GGCAATCCAAAGCTGAATCCAGCGTGTGTGCGCTATGGCGCGAGTCTCTCTGAGGAC 1914
 |||||
 Db 9852 GGCAATCCAAAGCTGAATCCAGCGTGTGTGCGCTATGGCGCGAGTCTCTCTGAGGAC 9911
 |||||
 QY 1915 TCTGACGAGGAGTCAAGGGGCTGGGGAAGAGATGGATTTGGTACGATTTGAGTCC 1974
 |||||
 Db 9912 TCTGACGAGGAGTCAAGGGGCTGGGGAAGAGATGGATTTGGTACGATTTGAGTCC 9971
 |||||
 QY 1975 CACTCTCCCGCCAAAGATGCCGAAGATGAGGAGGCCACTCCGCAAGAGTTCGAAGAACC 2034
 |||||
 Db 9972 CACTCTCCCGCCAAAGATGCCGAAGATGAGGAGGCCACTCCGCAAGAGTTCGAAGAACC 10031
 |||||
 QY 2035 ATGACCTTAAACGGTGTCTAATAGTGCACACAGCGACAGTGACCCCGAAAGAAACGCGCTA 2094
 |||||
 Db 10032 ATGACCTTAAACGGTGTCTAATAGTGCACACAGCGACAGTGACCCCGAAAGAAACGCGCTA 10091
 |||||
 QY 2095 GCGCTGATGTGCGAGCTGCCAGCCAGCTGCCCTGCACTCAGAAAAATCCCTTTGCT 2154
 |||||
 Db 10092 GCGCTGATGTGCGAGCTGCCAGCCAGCTGCCCTGCACTCAGAAAAATCCCTTTGCT 10151
 |||||
 QY 2155 AAGCAAAACGGTCTTCTCTGGAAGTTGA 2182
 |||||
 Db 10152 AAGCAAAACGGTCTTCTCTGGAAGTTGA 10179
 |||||

RESULT 11

AAK70011

ID AAK70011 standard; DNA; 21436 BP.

XX

AC AAK70011;

DT 06-NOV-2001 (first entry)

XX

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24823.

XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KW	cytostatic; gene therapy; vaccine; metastasis; ds.	PR	25-SEP-2000;	2000US-0234998P.
XX		PR	26-SEP-2000;	2000US-0235484P.
OS		PR	27-SEP-2000;	2000US-0235834P.
XX	Homo sapiens.	PR	27-SEP-2000;	2000US-0235836P.
XX		PR	29-SEP-2000;	2000US-0236327P.
PN	WO200157182-A2.	PR	29-SEP-2000;	2000US-0236367P.
XX		PR	29-SEP-2000;	2000US-0236368P.
XX		PR	29-SEP-2000;	2000US-0236369P.
PD	09-AUG-2001.	PR	29-SEP-2000;	2000US-0236370P.
XX		PR	02-OCT-2000;	2000US-0236802P.
XX		PR	02-OCT-2000;	2000US-0237037P.
XX		PR	02-OCT-2000;	2000US-0237038P.
PR		PR	02-OCT-2000;	2000US-0237039P.
PR		PR	02-OCT-2000;	2000US-0237040P.
PR		PR	13-OCT-2000;	2000US-0239935P.
PR		PR	13-OCT-2000;	2000US-0239937P.
PR		PR	20-OCT-2000;	2000US-0240960P.
PR		PR	20-OCT-2000;	2000US-0241221P.
PR		PR	20-OCT-2000;	2000US-0241785P.
PR		PR	20-OCT-2000;	2000US-0241787P.
PR		PR	20-OCT-2000;	2000US-0241808P.
PR		PR	20-OCT-2000;	2000US-0241809P.
PR		PR	20-OCT-2000;	2000US-0241826P.
PR		PR	01-NOV-2000;	2000US-024617P.
PR		PR	08-NOV-2000;	2000US-0246474P.
PR		PR	08-NOV-2000;	2000US-0246475P.
PR		PR	08-NOV-2000;	2000US-0246476P.
PR		PR	08-NOV-2000;	2000US-0246477P.
PR		PR	08-NOV-2000;	2000US-0246478P.
PR		PR	08-NOV-2000;	2000US-0246523P.
PR		PR	08-NOV-2000;	2000US-0246524P.
PR		PR	08-NOV-2000;	2000US-0246525P.
PR		PR	08-NOV-2000;	2000US-0246526P.
PR		PR	08-NOV-2000;	2000US-0246527P.
PR		PR	08-NOV-2000;	2000US-0246528P.
PR		PR	08-NOV-2000;	2000US-0246532P.
PR		PR	08-NOV-2000;	2000US-0246609P.
PR		PR	08-NOV-2000;	2000US-0246611P.
PR		PR	08-NOV-2000;	2000US-0246613P.
PR		PR	17-NOV-2000;	2000US-0249207P.
PR		PR	17-NOV-2000;	2000US-0249208P.
PR		PR	17-NOV-2000;	2000US-0249209P.
PR		PR	17-NOV-2000;	2000US-0249211P.
PR		PR	17-NOV-2000;	2000US-0249212P.
PR		PR	17-NOV-2000;	2000US-0249213P.
PR		PR	17-NOV-2000;	2000US-0249214P.
PR		PR	17-NOV-2000;	2000US-0249215P.
PR		PR	17-NOV-2000;	2000US-0249216P.
PR		PR	17-NOV-2000;	2000US-0249217P.
PR		PR	17-NOV-2000;	2000US-0249218P.
PR		PR	17-NOV-2000;	2000US-0249244P.
PR		PR	17-NOV-2000;	2000US-0249245P.
PR		PR	17-NOV-2000;	2000US-0249264P.
PR		PR	17-NOV-2000;	2000US-0249265P.
PR		PR	17-NOV-2000;	2000US-0249297P.
PR		PR	17-NOV-2000;	2000US-0249299P.
PR		PR	17-NOV-2000;	2000US-0249300P.
PR		PR	01-DEC-2000;	2000US-0250160P.
PR		PR	01-DEC-2000;	2000US-0250391P.
PR		PR	05-DEC-2000;	2000US-0251300P.
PR		PR	05-DEC-2000;	2000US-0251988P.
PR		PR	05-DEC-2000;	2000US-0256719P.
PR		PR	06-DEC-2000;	2000US-0251479P.
PR		PR	08-DEC-2000;	2000US-0251856P.
PR		PR	08-DEC-2000;	2000US-0251868P.
PR		PR	08-DEC-2000;	2000US-0251869P.
PR		PR	08-DEC-2000;	2000US-0251989P.
PR		PR	08-DEC-2000;	2000US-0251990P.
PR		PR	11-DEC-2000;	2000US-0254097P.
PR		PR	05-JAN-2001;	2001US-0259678P.

XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX Disclosure; SEQ ID NO 24923; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention
XX
XX Sequence 21436 BP; 5267 A; 4950 C; 5418 G; 5801 T; 0 U; 0 Other;
XX
XX Query Match 30.1%; Score 727.2; DB 4; Length 21436;
XX Best Local Similarity 98.3%; Pred. No. 1e-203;
XX Matches 735; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
XX
QY 1435 CAGCTTCCTCTCAGATGATAAAGATCCACTCACTTAAATGGGACTGGACATTGAAA 1494
DB 9432 CAGGTGCGCTCTGCTTACTTCCAGAAATCCACCTCACTTAAATGGGACTGGACATTGAAA 9491
QY 1495 GACAGCGCAAGCAGTCCATGTCGAGTCTTAACGGGAATTCAGTGTCAACAGGGCTAGT 1554
DB 9492 GACAGCGCAAGCAGTCCATGTCGAGTCTTAACGGGAATTCAGTGTCAACAGGGCTAGT 9551
QY 1555 CTTGTTAATGCTTCAGCTTCTGTCCAAACTGGTCAAGTAAATAGTCTCAGTGTATCCCA 1614
DB 9552 CTTGTTAATGCTTCAGCTTCTGTCCAAACTGGTCAAGTAAATAGTCTCAGTGTATCCCA 9611
QY 1615 GAAATCTTAAGAAACAAAATAATTAATCAGTATTCACAAAGTTGCTGTTCGCCAG 1674
DB 9612 GAAATCTTAAGAAACAAAATAATTAATCAGTATTCACAAAGTTGCTGTTCGCCAG 9671
QY 1675 TGTGAGTCTCAACCTTAACCTTCATAGTAATTTCTTGGAGAACCTTACCAAGCCGTTCC 1734
DB 9672 TGTGAGTCTCAACCTTAACCTTCATAGTAATTTCTTGGAGAACCTTACCAAGCCGTTCC 9731
QY 1735 TCTTCTACCTTACCAATTTCTGAGTACAGTCTACCTCGAAGCATCTACCATCTCAGTT 1794
DB 9732 TCTTCTACCTTACCAATTTCTGAGTACAGTCTACCTCGAAGCATCTACCATCTCAGTT 9791
QY 1795 TCTAGTAAAGTAACAAACCGATCCCGCGAGTGAATCTCTCCAGCCCGTGTATGAAT 1854
DB 9792 TCTAGTAAAGTAACAAACCGATCCCGCGAGTGAATCTCTCCAGCCCGTGTATGAAT 9851
QY 1855 GGCAAAATCCAGCTGAATCCAGGCTGCTGGTGGCCCTATGGCGCGAGTCTCTTGAGGAC 1914
DB 9852 GGCAAAATCCAGCTGAATCCAGGCTGCTGGTGGCCCTATGGCGCGAGTCTCTTGAGGAC 9911
QY 1915 TCTGACGAGGAGTCAAGGGGCTGGGCAAGAGAAATGGGATTTGGTACGATTTGAGCTCC 1974
DB 9912 TCTGACGAGGAGTCAAGGGGCTGGGCAAGAGAAATGGGATTTGGTACGATTTGAGCTCC 9971
QY 1975 CACTCTCCCGCCCAAGATGCCGAAGATGAGGAGGCCCACTCCGACGAGCTTCAAGAACCC 2034

DB 9972 CACTCTCCCGCCCAAGATGCCGAAGATGAGGAGGCCCACTCCGACGAGCTTCAAGAACCC 10031
QY 2035 ATGACCCCTAAACGGTGTCTAATAGTGCAGACAGGACAGTGCACCGAAGAAACGGCTTA 2094
DB 10032 ATGACCCCTAAACGGTGTCTAATAGTGCAGACAGGACAGTGCACCGAAGAAACGGCTTA 10091
QY 2095 GCGCCTGATGTCGCCAGCTGCCAAGGCCAGCTGCCCTGCACCTCAGAAAAATCCCTTTGCT 2154
DB 10092 GCGCCTGATGTCGCCAGCTGCCAAGGCCAGCTGCCCTGCACCTCAGAAAAATCCCTTTGCT 10151
QY 2155 AAGGCAAAACGGTCTTCTCTGAAAAAGTTGA 2182
DB 10152 AAGGCAAAACGGTCTTCTCTGAAAAAGTTGA 10179
RESULT 12
AAS41077
ID AAS41077 standard; cDNA; 1122 BP.
XX AAS41077;
XX
XX 17-DEC-2001 (first entry)
XX
XX cDNA encoding novel human enzyme polypeptide #293.
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
XX ligase; hyperproliferative disorder; immunodeficiency disorder;
XX autoimmune disorder; neurological disorder; metabolic disorder;
XX inflammatory disorder; cardiovascular disorder; reproductive disorder;
XX blood-related disorder; infectious disorder; gene therapy; cytostatic;
XX anti arthritic; nephrotropic; anticoagulant; ss.
XX Homo sapiens.
XX
XX WO200155301-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001239.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225367P.
XX 14-AUG-2000; 2000US-0225368P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.

```
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0228287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249266P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249298P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0234097P.
PR 05-JAN-2001; 2001US-0239678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-465566/50.
XX P-PSDB; AAU23207.
XX Novel polypeptides and polynucleotides useful for diagnosing, preventing,
XX treating neural, immune system, muscular, reproductive, pulmonary,
XX cardiovascular, renal, proliferative disorders and cancerous diseases.
XX Claim 4; SEQ ID NO 303; 1180pp; English.
XX The present invention relates to the isolation of novel human enzyme
XX polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
XX encoding them. The enzyme polypeptides of the invention may comprise the
XX functional classes of oxidoreductases, transferases, hydrolases, lyases,
XX isomerases or ligases. The sequences of the invention are useful in the
XX diagnosis, treatment, prevention and/or prognosis of a wide range of
XX disorders including hyperproliferative disorders (e.g. cancer),
XX immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.
XX arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic
XX disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),
XX cardiovascular disorders (e.g. atherosclerosis), blood-related disorders
XX (e.g. haemophilia), reproductive disorders (e.g. infertility) and
XX infectious disorders (e.g. influenza). The polynucleotides of the
XX invention can also be used in gene therapy. AA540785-AA541684 represent
XX cDNA sequences encoding for the novel human enzyme polypeptides of the
XX invention. Note: The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 1122 BP; 330 A; 247 C; 260 G; 285 T; 0 U; 0 Other;
SQ
Query Match 21.5%; Score 517.4; DB 4; Length 1122;
Best Local Similarity 97.1%; Pred. No. 3.1e-142;
Matches 527; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 1 GGGCTGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 60
DB 12 GGGCTGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 71
```

QY	61	CAGAGTTGAACAATGACCATAGTATTGACAAAGCTTCTGAAATCTTTCAGACCCATCAGCCTAT	120
DB	72	CAGAGTTGAACAATGACCATAGTATTGACAAAGCTTCTGAAATCTTTCAGACCCATCAGCCTAT	131
QY	121	CAGAAATCAGCCTGGCAGCTCCGAGGCGAGTCTCACTTGGAGACATGATGACGGTTCTTGCC	180
DB	132	CAGAAATCAGCCTGGCAGCTCCGAGGCGAGTCTCACTTGGAGACATGATGACGGTTCTTGCC	191
QY	181	AGCTGGGGTGTGTGTCTTCATTGAATGATGTCTCAAAATCACACACTTTCTTTAGGACCA	240
DB	192	AGCTGGGGTGTGTGTCTTCATTGAATGATGTCTCAAAATCACACACTTTCTTTAGGACCA	251
QY	241	GTAACCTGGTGTGTAGTATTATTCGAGTTTCATCTGTACCTGATAAATCAAAACCATCACCA	300
DB	252	GTAACCTGGTGTGTAGTATTATTCGAGTTTCATCTGTACCTGATAAATCAAAACCATCACCA	311
QY	301	CAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCTCTCCACAGAAAGTTCTTTTCCCATCT	360
DB	312	CAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCTCTCCACAGAAAGTTCTTTTCCCATCT	371
QY	361	GAGAAGATTGTGTTTAAAGTGGCAACAAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTG	420
DB	372	GAGAAGATTGTGTTTAAAGTGGCAACAAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTG	431
QY	421	GCGAATACCTGTTTGGCCAAATCGAGCACTGCAGTGTTTAACTACACACACCTTTGCCC	480
DB	432	GCGAATACCTGTTTGGCCAAATCGAGCACTGCAGTGTTTAACTACACACACCTTTGCCC	491
QY	481	AATTACATGCTATACATGACACACCTCCAAACATGTCAATGAGAGGCTTTTGTATGATG	540
DB	492	AATTACATGCTATACATGAGAACACTCCAAACATGTCAATGAGAGGCTTTTGTATGATG	551
QY	541	TGT 543	
DB	552	TAT 554	
RESULT 13			
AAS41613			
ID	AAS41613 standard; cDNA; 577 BP.		
XX	AAS41613;		
AC			
XX			
DT	17-DEC-2001 (first entry)		
XX			
DE	cDNA encoding novel human enzyme polypeptide #829.		
XX			
KW	Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase; ligase; hyperproliferative disorder; immunodeficiency disorder; autoimmune disorder; neurological disorder; metabolic disorder; inflammatory disorder; cardiovascular disorder; reproductive disorder; blood-related disorder; infectious disorder; gene therapy; cytostatic; anti arthritic; nephrotropic; anticoagulant; ss.		
OS	Homo sapiens.		
XX			
PN	WO200155301-A2.		
XX			
PD	02-AUG-2001.		
XX			
PF	17-JAN-2001; 2001WO-US001239.		
XX			
PR	31-JAN-2000; 2000US-0179065P.		
PR	04-FEB-2000; 2000US-0180628P.		
PR	24-FEB-2000; 2000US-0184664P.		
PR	02-MAR-2000; 2000US-0186350P.		
PR	16-MAR-2000; 2000US-0189874P.		
PR	17-MAR-2000; 2000US-0190076P.		
PR	18-APR-2000; 2000US-0198123P.		
PR	19-MAY-2000; 2000US-0205515P.		
PR	07-JUN-2000; 2000US-0209467P.		
PR	28-JUN-2000; 2000US-0214886P.		

PR	20-OCT-2000;	2000US-0241787P.
PR	20-OCT-2000;	2000US-0241808P.
PR	20-OCT-2000;	2000US-0241809P.
PR	20-OCT-2000;	2000US-0241826P.
PR	01-NOV-2000;	2000US-0244617P.
PR	08-NOV-2000;	2000US-0246474P.
PR	08-NOV-2000;	2000US-0246475P.
PR	08-NOV-2000;	2000US-0246476P.
PR	08-NOV-2000;	2000US-0246477P.
PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246609P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246611P.
PR	08-NOV-2000;	2000US-0246613P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249246P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-02559678P.
XX		
PA	(HUMA-)	HUMAN GENOME SCI INC.
XX		
PI	Rosen CA,	Barash SC, Ruben SM;
XX		
DR	WPI:	2001-4655566/50.
DR	P-PSDB;	AAU23743.
XX		
PT	Novel polypeptides and polynucleotides useful for diagnosing, preventing,	
PT	treating neural, immune system, muscular, reproductive, pulmonary,	
PT	cardiovascular, renal, proliferative disorders and cancerous diseases.	
XX	Claim 4;	SEQ ID NO 839; 1180pp; English.
PS		
XX	The present invention relates to the isolation of novel human enzyme	
CC	polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences	
CC	encoding them. The enzyme polypeptides of the invention may comprise the	
CC	functional classes of oxidoreductases, transferases, hydrolases, lyases,	
CC	isomerases or ligases. The sequences of the invention are useful in the	
CC	diagnosis, treatment, prevention and/or prognosis of a wide range of	

FN US6331427-B1.
 PD 18-DEC-2001.
 XX 26-MAR-1999; 99US-00280116.
 PF 26-MAR-1999; 99US-00280116.
 PR (MILL-) MILLENNIUM PHARM INC.
 PA Robison KE;
 XX WPI; 2002-129545/17.
 DR New polynucleotides encoding protease homologs of the G-protein-coupled
 PT protease family, useful in identifying agonists and antagonists for
 PT diagnosis and treatment of protease mediated disorders.
 XX Disclosure; Col 251-252; 245pp; English.
 CC The invention relates to an isolated human protease nucleic acid molecule
 CC comprising a nucleotide sequence of 546 base pairs, one of 268 fully
 CC defined in the specification. Also disclosed are production of an
 CC isolated polypeptide encoded by the nucleic acid, comprising introducing
 CC the nucleic acid into a host cell and culturing under conditions to
 CC express the protein from the nucleic acid, use of an antibody to detect
 CC the encoded protein in a sample and to modulate its in vivo activity,
 CC identifying agents that bind to the protein and identification of a
 CC polynucleotide agent that modulates the expression of the nucleic acid or
 CC its complement (i.e. gene therapy). The nucleic acid can be used to
 CC identify an agent that modulates the expression of the nucleic acid or
 CC nucleic acid, and can be used to isolate the protein. The nucleic acid
 CC can be used in diagnostic assays for determining nucleic acid expression
 CC as well as activity in the context of a biological sample (e.g., blood,
 CC serum, cells, tissue) to determine whether an individual has a disease or
 CC disorder, or is at risk of developing a disease or disorder, associated
 CC with aberrant expression or activity of the nucleic acid. The nucleic
 CC acid can be used to detect mutations in protease genes and gene
 CC expression products such as mRNA. The nucleic acid can be used as
 CC hybridisation probes to detect naturally-occurring genetic mutations in a
 CC protease gene. The nucleic acid can be used in drug screening methods to
 CC identify agonists and antagonists that can be used to diagnose and treat
 CC such protease mediated disorders e.g., proliferative, differentiative,
 CC developmental or haematopoietic disorders. The nucleic acid can be used
 CC as probes, primers, in biological assays, to determine patterns of gene
 CC expression, to design ribozymes and to construct transgenic animals. The
 CC present sequence represents one of the 268 disclosed human G-protein-
 CC coupled protease cDNA sequences
 XX
 SQ Sequence 722 BP; 189 A; 186 C; 145 G; 196 T; 0 U; 6 Other;
 Query Match 19.7%; Score 474.6; DB 6; Length 722;
 Best Local Similarity 92.9%; Pred. No. 1.2e-129;
 Matches 562; Conservative 0; Mismatches 33; Indels 10; Gaps 6;
 QY 1061 AGCATGTGAATACCTGAGTACTTCTGATATCGGCATATATGCTCAACCCACGGAG 1120
 DB 107 AGATGTGAATACCTGAGTACTTCTGATATCGGCATATATGCTCAACCCACGGAG 166
 QY 1121 AGCAATTGTCTACGCTTCTGATGCGTGTGCTCCACACTGTTTAAATGCCATGCTG 1180
 DB 167 AGCAATTGTCTACGCTTCTGATGCGTGTGCTCCACACTGTTTAAATGCCATGCTG 226
 QY 1181 GCATTACTTCTGCTACATAAAAGCTAGCAATCGCTCTGGTATCAATGAATGACTCCA 1240
 DB 227 GCATTACTTCTGCTACATAAAAGCTAGCAATCGCTCTGGTATCAATGAATGACTCCA 286
 QY 1241 TTGTATCTACAGTATATAGATCGGTACTCAGCCAAAGCTATGCTCTTTTATA 1300
 DB 287 TTGTATCTACAGTATATAGATCGGTACTCAGCCAAAGCTATGCTCTTTTATA 346
 QY 1301 TCAGGTCCCATCATGTGAAATATGAGGTGAAGTACTCATCCACCCATGCCCGGCC 1360
 DB 347 TCAGTCCCATCATGTGAAATATGAGGTGAAGTACTCATCCACCCATGCCCGGCC 406
 QY 1361 AGTCTCTCTCCCGCCCGTCACTAGTCAGCGGGTGTCCACAAACAAAGGCTGCCAG 1420
 DB 407 AGTCTCTCTCCCGCCCGTCACTAGTCAGCGGGTGTCCACAAACAAAGGCTGCCAG 465
 QY 1421 GCTTTATCGGACCAACAGCTTCCCTCTCATCATGATAAAGAAATCCACCTCACTTAATGGGA 1480
 DB 466 GCTTTATCGGACCAACAGCTTCCCTCTCATCATGATAAAGAAATCCACCTCACTTAATGGGA 525
 QY 1481 C-TGGACCATTTGAAGACAGCCCAAGCAGTTCCATGTCGAGTCCCTAACGGGAATTCAGT 1539
 DB 526 CATGGACCATTTGAAGACAGCCCAAGCAGTTCCATGTCGAGTCCCTAACGGGAATTCAGT 585
 QY 1540 GTCAACAGGGCTAGTCTCTGTTAATGCTTCACTTCTGTCCTCAAACTCGTCAGTTAATAGG 1599
 DB 586 GTCAACAGGGCTAG-CTGTGTTAATGTT-CTTGTGTCCTCAAAATGCTCAGTTAATAGG 640
 QY 1600 TCTCTAGTGTATCCAGAACATCTTAAGAAACAAAAAATTACAATCAGTATTCACAAAG 1659
 DB 641 T-CTCAGTGTATCC-AGAACATCTTANGAACAAAAAATTCAANAGTATNACAACAAGTTG 697
 QY 1660 TTGCC 1664
 DB 698 CTGTC 702
 RESULT 15
 AAH05659
 ID AAH05659 standard; cDNA; 521 BP.
 AC AAH05659;
 XX 26-JUN-2001 (first entry)
 DT Human cDNA clone (5'-primer) SEQ ID NO:2494.
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 KW Homo sapiens.
 OS Homo sapiens.
 PN EP1074617-A2.
 XX 07-FEB-2001.
 PD 28-JUL-2000; 2000EP-00116126.
 PF 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX (HELI-) HELIX RES INST.
 PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 DR Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.
 PS Claim 1; SEQ ID NO 2494; 2537pp + Sequence Listing; English.
 XX The present invention describes primer sets for synthesising 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

Search completed: August 16, 2004, 13:39:37
Job time : 1360 secs

Result No.	Score	Query t		DB	ID	Description
		Match	Length			
1	1148.2	47.6	1485	11	AK010801	Mus muscu
2	986.2	40.9	1201	13	BX418055	BX418055
3	942.8	39.1	1063	12	BM920091	BM920091
4	910.2	37.7	1144	13	BX331803	BX331803

11042159

AK010801 1485 bp mRNA linear HTC 20-SEP-2003
Mus musculus ES cells cDNA, RIKEN full-length enriched library,
clone:241014K03 product:hypothetical Ubiquitin carboxyl-terminal
hydrolase family 2 containing protein, full insert sequence.
Accession number: AK010801

D b	964	ATGTCACGCCAATGGAGAGCCAATTAATTTATGTTTGTATGCTGCTGGTGCACT	1023
Q y	1162	GGTTTAAATTGGCATGCTGGGCATTACTTCTGCTACATAAAGCTAGCAATGGCCTCTGG	1221
D b	1024	GGTTTAAATTGTCATGCTGGGCCACTACTTTTGCTACATCAAGGTAGCAATGGCCTCTGG	1083
Q y	1222	TATCAAAATGAATGACTCAATTGTTATCTACCAAGTCATATTAGATCGGTACTCAGCCAA	1281
D b	1084	TATCAGATGAATGACTCAATGCTGCCACAGTGATATCAGAGCAGTCTTAAACGCAA	1143
Q y	1282	GCCTATGTGCTCTTTTATATCAGTGTCCATGATGTGAAATGGAGTGAACTTCTCAT	1341
D b	1144	GCTTACGTGCTCTTTTATATCAGTCCCATGTGTAATGAAATGGAGGTAGTCTGCTCAT	1203
Q y	1342	CCACCCCATAGCCCCGGCGAGTCTCTCCCGGCCCGTCACTAGTCAGCGGGTTGTCA	1401
D b	1204	CTTGCCCATAGCCCCGGCCCAATCTCTCCCGGCCAGAGTCAGTCAGCGGGTAGTCA	1263
Q y	1402	AACAAACAGGCTGCGCCAGGCTTTATCGGACCACAGCTTCCCTCTCACATGATAAGAA	1461
D b	1264	AACAAGCAGGTGGTCCAGGGTTTATTGGACCCCAAGCTGCCATGTGATGAGA	1323
Q y	1462	CCACCTCACATTAATGGAGTGAACCATTAAGAACGCCAAGCAGTTCATGTGAGT	1521
D b	1324	ATGGCACATTTGAATGGCACCGCCAGTGAAGACACACCAGGTAGTTCTGTGTCA	1383
Q y	1522	CCTAACGGGAATTCAGTGTCAACAGGGTAGTCTCTGTTAATGCTTCAGCTTCTG	1581
D b	1384	CCTAACGGAAACACCAAGCGTCAATAGGGCCAGTCTCTGCTACTGCTTCGAGT	1443
Q y	1582	AACGTGTCAAGTTAATAGTTCCTCAGTGATCCCAAGATCC	1622
D b	1444	AACTGGTCTGTTACCAAGACCCCTCAGTTATTCAGATCA	1484

RESULT 2					
BX418055					
LOCUS	BX418055	1201 bp	mRNA	linear	EST 13-MAY-2003
DEFINITION	Homo sapiens FETAL BRAIN Homo sapiens cDNA clone				
ACCESSION	CS0DF001YG23 5-PRIME, mRNA sequence.				

BK418035.1 GI:30646403
 EST.
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1201)
 Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)

Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 4596.f For
 more information about this cluster, see
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSODF001AD12QP2&cluster=4596.f>. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CSODF001AD12QP2.

[illegible]

```
1. 1201
/organisms="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF001Y23"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens"
/note="Organ: brain; Ver:
```

was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match	40.9%;	Score 986.2;	DB 13;	Length 1201;
Best Local Similarity	98.6%;	Pred. No. 1.6e-253;		
Matches 1023				

Accession	10221	Conservative	4	Mismatches	7	Indels	4	Gaps
QY	17	CGGCGGCGGCGCGAGGGGGATGGAGCGCGCGAGCGGGT	CAGATTGAA	CAATGA	76			
Db	60	CGGGATCGGCGCGGAGGGGATGGAGCGCGCGAGCGGGT	CAGATTGAA	CAATGA	119			
QY	77	CCATAGTTGACAAAGCTTCTGAAATCTTCAGACCCATCAGCCTATCAGAAATCAGCTGGCA	136					
Db	120	CCATAGTTGACAAAGCTTCTGAAATCTTCAGACCCATCAGCCTATCAGAAATCAGCTGGCA	179					
QY	137	GTCCGAGGCGAGTCCACCTGGAGACATGGATGCAAGTTTGCACAGCTGGGTTGCTGT	196					
Db	180	GTCCGAGGCGAGTCCACCTGGAGACATGGATGCAAGTTTGCACAGCTGGGTTGCTGT	239					
QY	197	CTTCATTGAATGATGTGTCAAATTCACACTTCTTTAGGACCCAGTACCTGGTGTGAG	256					
Db	240	CTTCATTGAATGATGTGTCAAATTCACACTTCTTTAGGACCCAGTACCTGGTGTGAG	299					
QY	257	TTTATTCGAGTTCATCTGPACTGATAAATCAAAACCATCACCAAAAGGATCAAGCCC	316					
Db	300	TTTATTCGAGTTCATCTGPACTGATAAATCAAAACCATCACCAAAAGGATCAAGCCC	359					
QY	317	TAGGTGATGGCATCGCTCCTCCACAGAAAGTTCTTTCCCATCGAGAAATTTGCTT	376					
Db	360	TAGGTGATGGCATCGCTCCTCCACAGAAAGTTCTTTCCCATCGAGAAATTTGCTT	419					
QY	377	AGTGGCAACAAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTT	436					
Db	420	AGTGGCAACAAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTT	479					
QY	437	CCATGCGACGCTGAGTGTAACTACACACACACCTCTTGCCAAATTCATGCTATCAC	496					
Db	480	CCATGCGACGCTGAGTGTAACTACACACACCTCTTGCCAAATTCATGCTATCAC	539					
QY	497	ATGAACACTCCAAACATGTCGACAGAGCTTTTGTATGTTGTACAATGCAAGCAC	556					
Db	540	ATGAACACTCCAAACATGTCGACAGAGCTTTTGTATGTTGTACAATGCAAGCAC	599					
QY	557	ATATTACCCAGGCACCTCAGTAATCCTGGGACGTTATTAACCAATGTTTGTATCAATG	616					
Db	600	ATATTACCCAGGCACCTCAGTAATCCTGGGACGTTATTAACCAATGTTTGTATCAATG	659					
QY	617	AGATGCGGCGTATAGCTAGGCACTCCGTTTGGAAACCAAGAGATGCCATGAATCC	676					
Db	660	AGATGCGGCGTATAGCTAGGCACTCCGTTTGGAAACCAAGAGATGCCATGAATCC	719					
QY	677	TTCAATACACTGTTGATGCTATGACAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACA	736					
Db	720	TTCAATACACTGTTGATGCTATGACAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACA	779					
QY	737	GACACACCCAGGCCACACTCTTGTGTCAGATATTTGGAGGATACCTAGATCTAGAG	796					
Db	780	GACACACCCAGGCCACACTCTTGTGTCAGATATTTGGAGGATACCTAGATCTAGAG	839					
QY	797	TCAATTTGTTAAATGCAAGCGGTTTCAGATACCTTTTGATCCATATCTTCATATAACAT	856					
Db	840	TCAATTTGTTAAATGCAAGCGGTTTCAGATACCTTTTGATCCATATCTTCATATAACAT	899					
QY	857	TGGAGATAAAGCTGCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTGAAGCCGGAAC	916					
Db	900	TGGAGATAAAGCTGCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTGAAGCCGGAAC	959					
QY	917	AGCTTGATGGAGAAAACCTGCTACAGTGCACAGTGTAAAGATGGTTCAGCTTCAA	976					
Db	960	AGCTTGATGGAGAAAACCTGCTACAGTGCACAGTGTAAAGATGGTTCAGCTTCAA	1019					

Wed Aug 18 13:53:02 2004

977 AGAGGTTCACTATCCATAGATCCCTCTAATGTTCTTACACATTTCTCTGAAACGTTTGGCAA 1036
 1020 AGAGG-TYACTATCATAGATCCCTCTAATGTTCTTACAC-TTCTCTGAAAGGTTT-GCA 1075
 1037 ATTTTACCGGTGGA AAA 1053
 1076 AATTACCGGKGAAAA 1092

RESULT 3
 BM920091 1063 bp mRNA linear EST 12-MAR-2002
 AGENCOURT_6706520 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5749963
 5', mRNA sequence.

BM920091
 BM920091.1 GI:19370470
 EST.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 AUTHORS NIH-MGC http://img.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapps-f@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM12780 row: c column: 20
 High quality sequence stop: 760.

FEATURES
 source
 1..1063
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5749963"
 /lab_host="DH10B"
 /clone_lib="NIH MGC 120"
 /note="Organ: pooled pancreas and spleen; Vector:
 pCMV-SPORT6; Site: 1: NotI; Site 2: EcoRV (destroyed); RNA
 source anonymous pool of spleen and pancreas from 28 yo
 male. Library is oligo-dT primed and directionally cloned
 (EcoRV site is destroyed upon cloning). Average insert
 size 1.5 kb, insert size range 1-2.5 kb. Library is
 normalized and enriched for full-length clones and was
 constructed by C. Gruber (Invitrogen). Research Genetics
 tracking code 025. Note: this is a NIH_MGC Library."

ORIGIN
 Query Match 39.1%; Score 942.8; DB 12; Length 1063;
 Best Local Similarity 96.6%; Pred. No. 7e-242;
 Matches 994; Conservative 0; Mismatches 30; Indels 5; Gaps 3;

513 ATGTCATGAGAGGCTTTTGTATGATGTGTACATGCAAGACATATTACCCAGGCACT 572
 17 ATGTCATGAGAGGCTTTTGTATGATGTGTACATGCAAGACATATTACCCAGGCACT 76
 573 CAGTAATCTCGGGAGGCTTTTGTATGATGTGTACATGCAAGACATATTACCCAGGCACT 632
 77 CAGTAATCTCGGGAGGCTTTTGTATGATGTGTACATGCAAGACATATTACCCAGGCACT 136
 633 TAGGCACCTCCGTTTGGAAACCAAGAGATGCCCATGAATCTCTCAATACACTGTGA 692
 137 TAGGCACCTCCGTTTGGAAACCAAGAGATGCCCATGAATCTCTCAATACACTGTGA 196
 693 TGCTATGAGAGGAGCATGCTTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCGAC 752

197 TGCTATGAGAGGAGCATGCTTTGAATGGCAGCAATAAATTAGACAGACACCCAGGCGAC 256
 753 CACTCTTGTGTTGTGAGATATTTGGAGGATACCTAAGATCTTAGAGTCAAAATGTTAAATTG 812
 257 CACTCTTGTGTTGTGAGATATTTGGAGGATACCTAAGATCTTAGAGTCAAAATGTTAAATTG 316
 813 CAAGGGCGTTTCAGANTACTTTTGATCCATATCTTTGATATAAATTGGAGATAAAGGCTGC 872
 317 CAAGGGCGTTTCAGANTACTTTTGATCCATATCTTTGATATAAATTGGAGATAAAGGCTGC 376
 873 TCAGAGTGTCAAGAGGCAATTTGGAGGATTTGAGAGCGGAAACAGCTTGTATGGAGAAA 932
 377 TCAGAGTGTCAAGAGGCAATTTGGAGGATTTGAGAGCGGAAACAGCTTGTATGGAGAAA 436
 933 CTCGTACAGAGTGCAACAAAGTGTAAAGAGATGTTCCAGCTTCAAGAGGTTCACTATCCA 992
 437 CTCGTACAGAGTGCAACAAAGTGTAAAGAGATGTTCCAGCTTCAAGAGGTTCACTATCCA 496
 993 TAGATCCTCTAATGTTCTTACACTTTCTGAAACGTTTGGCAAAATTTTACCGGTGAAA 1052
 497 TAGATCCTCTAATGTTCTTACACTTTCTGAAACGTTTGGCAAAATTTTACCGGTGAAA 556
 1053 AATTGCTAAGAGTGTGAATACCTGAGTATCTTGATATTCGGCCATATATGTTCTCAACC 1112
 557 AATTGCTAAGAGTGTGAATACCTGAGTATCTTGATATTCGGCCATATATGTTCTCAACC 616
 1113 CAAGGAGAGGCAATTTGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1172
 617 CAAGGAGAGGCAATTTGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 676
 1173 CCATGCTGGCCATTTCTGCTACATAAAGTAGCAATGGCTCTGTTATCAAAATGAA 1232
 677 CCATGCTGGCCATTTCTGCTACATAAAGTAGCAATGGCTCTGTTATCAAAATGAA 736
 1233 TGACTCCATTTGTTCTTACAGTGTATTTAGATCGGTACTCAGCCAAACAGCTATGCT 1292
 737 TGACTCCATTTGTTCTTACAGTGTATTTAGATCGGTACTCAGCCAAACAGCTATGCT 796
 1293 CTTTATATCAGGTCCCATGATGTGAAATAATGGAGGTGAATTTACTATCCACCCATAG 1352
 797 CTTTATATCAGGTCCCATGATGTGAAATAATGGAGGTGAATTTACTATCCACCCATAG 856
 1353 CCCCAGGAGGTC-CTCTCCCGCCCGCTCATCAGTCAAGGGGTTGTACCAACAAACAGG 1411
 857 CCCCAGGAGGTC-CTCTCCCGCCCGCTCATCAGTCAAGGGGTTGTACCAACAAACAGG 916
 1412 CTGCGCCAGGCTTTATCGGACCAAG-CTTCCCTCTCATGATAAAGATCCACCTCAC 1470
 917 CTGCGCCAGGCTTTATCGGACCAAG-CTTCCCTCTCATGATAAAGATCCACCTCAC 976
 1471 TTAAA---TGGGACTGGACCATTTGAAGACAGCCCAAGAGTTCCATGTCGAGTCTTAAC 1527
 977 CTTAAATGGGAGTGGGACCATTTGAAGAGCCCGCCAGGAGTTTCCATGTCGAGGCTACC 1036
 1528 GGGAAATTC 1536
 1037 GGGAAATTC 1045

RESULT 4
 BX331803 1144 bp mRNA linear EST 02-MAY-2003
 LOCUS BX331803 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
 DEFINITION cDNA clone CS00C002YH20 5-PRIME, mRNA sequence.
 ACCESSION BX331803
 VERSION BX331803.1 GI:30339041
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1144)

AUTHORS		Li,W.B., Gruber,C., Jessee,J. and Polayes,D.	
TITLE		Full-length cDNA libraries and normalization	
JOURNAL		Unpublished (2001)	
COMMENT		Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 4596.f For more information about this cluster, see http://www.genoscope.cns.fr/ cgi-bin/cluster.cgi?seq=CS0DC002DD10QPl&cluster=4596.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DC002DD10QPl.	
FEATURES		Location/Qualifiers 1..1144 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CS0DC002YH20" /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED" /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED" /notes="1st strand cDNA was primed with a NotI-oligo (dr) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."	
source			
ORIGIN		Query Match 37.7%; Score 910.2; DB 13; Length 1144; Best Local Similarity 97.7%; Pred. No. 4.3e-233; Matches 927; Conservative 15; Mismatches 5; Indels 2; Gaps 2;	
QY	691	GATGCTATGCAGAAAGCATGCTTGAATCGCAGCAATAAATTAGACACACACCCAGGCC	750
DB	64	GATGCTATGCAGAAAGCATGCTTGAATCGCAGCAATAAATTAGACACACACCCAGGCC	123
QY	751	ACCACTCTTGTGTGCAGATATTTGGAGATACCTTAAGCTAGCTAGAGTCAATGTTAAAT	810
DB	124	ACCACTCTTGTGTGCAGATATTTGGAGATACCTTAAGCTAGAGTCAATGTTAAAT	183
QY	811	TGCAAGGCGTTTCAGATACCTTTTGAATCCATATCTTGATATACATTTGGAGATAAAGGCT	870
DB	184	TGCAAGGCGTTTCAGATACCTTTTGAATCCATATCTTGATATACATTTGGAGATAAAGGCT	243
QY	871	GCTCAGAGTGTCAACAGGATTTGGAGATACCTTAAGCTAGCTAGAGTCAATGTTAAAT	930
DB	244	GCTCAGAGTGTCAACAGGATTTGGAGATACCTTAAGCTAGCTAGAGTCAATGTTAAAT	303
QY	931	AATCGGTACAAGTGCAGCAAGTGTAAAGATGTTTCCAGCTTCAAGAGGTTCACTATC	990
DB	304	AATCGGTACAAGTGCAGCAAGTGTAAAGATGTTTCCAGCTTCAAGAGGTTCACTATC	363
QY	991	CATAGATCTCTTAATGTTCTTACACTTCTCTGAAACGTTTTCGAAATTTACCGGTGA	1050
DB	364	CATAGATCTCTTAATGTTCTTACACTTCTCTGAAACGTTTTCGAAATTTACCGGTGA	423
QY	1051	AAAATTGCTAAGCATGTGAAATACCTCGATATCTTGATATTCGGCCATPATGTCTCAA	1110
DB	424	AAAATTGCTAAGCATGTGAAATACCTCGATATCTTGATATTCGGCCATPATGTCTCAA	483
QY	1111	CCCAACGGAGAGCAATTGTCTACGTTCTGTATGTCAGTCTCGTCCACACTGGTTTAAAT	1170
DB	484	CCCAACGGAGAGCAATTGTCTACGTTCTGTATGTCAGTCTCGTCCACACTGGTTTAAAT	543
QY	1171	TGCCATGCTGGCCATTAATCTTGCTACATATAAAGCTAGCAATGGCTCTGGTATCAATG	1230
DB	544	TGCCATGCTGGCCATTAATCTTGCTACATATAAAGCTAGCAATGGCTCTGGTATCAATG	603
QY	1231	AATGACTCCATTGTATCTACCAAGTATATAGATCGGTACTCAGCCAAACAGCCCTATGTG	1290
DB	604	ARTGACTCCATTGTATCTACCAAGTATATAGATCGGTACTCAGCCAAACAGCCCTATGTG	663

QY 557 ATATTACCCAGGCACTCAGTAATCTCTGGGACGTTATTAAACCAATCTTTGTCTCATCAATG 616
 |||||
 Db 600 ATATTACCCAGGCACTCAGTAATCTCTGGGACGTTATTAAACCAATCTTTGTCTCATCAATG 659
 |||||
 QY 617 AGATGCGGCGTATAGCTAGGCACTCTCGTTTGGAAACCAAGAAGATGCCCATGAATTC 676
 |||||
 Db 660 AGATGCGGCGTATAGCTAGGCACTCTCGTTTGGAAACCAAGAAGATGCCCATGAATTC 719
 |||||
 QY 677 TTCAATACACTGTTGATGCTATCAGAAAGCATGCTTGAATGGCAGCAATAAATATAGACA 736
 |||||
 Db 720 TTCAATACACTGTTGATGCTATCAGAAAGCATGCTTGAATGGCAGCAATAAATATAGACA 779
 |||||
 QY 737 GACACACCCAGGCCACCACTCTTGTGTGATATTTGGAGGATACCTAAGATCTAGAG 796
 |||||
 Db 780 GACACACCCAGGCCACCACTCTTGTGTGATATTTGGAGGATACCTAAGATCTAGAG 839
 |||||
 QY 797 TCAATATGTTTAATTCAGAGGGGCTTTCAGATCTTTTGTATCCATATCTTGATATAACAT 856
 |||||
 Db 840 TCAATATGTTTAATTCAGAGGGGCTTTCAGATCTTTTGTATCCATATCTTGATATAACAT 899
 |||||
 QY 857 TGGAGATAAAGGCTGCTCAGAGTGTCACCAAGGCATTGGAGCAGTTTGTGAAG 909
 |||||
 Db 900 TGGAGATAAAGGCTGCTCAGAGTGTCACCAAGGCATTGGAGCAGTTTGTGAAG 952
 |||||

RESULT 7
 LOCUS BX451363 918 bp mRNA linear EST 22-MAY-2003
 DEFINITION BX451363 Homo sapiens NEUROBLASTOMA Homo sapiens cdna clone
 CLOBB028ZF10 5-PRIME, mRNA sequence.
 ACCESSION BX451363
 VERSION BX451363.1 GI:31034264
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 JOURNAL Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 COMMENT Full-length cDNA libraries and normalization
 Unpublished (2001)
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 5254.r For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0BAF001ZF08.AF00063.1&cluster=5254.r.
 Contact : Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0BAF001ZF08.AF00063.1.

FEATURES
 source
 1. 918
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CLOBB028ZF10"
 /tissue_type="NEUROBLASTOMA"
 /clone_lib="Homo sapiens NEUROBLASTOMA"
 /notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
 with a NotI-oligo (dT) primer. Five prime end enriched,
 double-strand cDNA was digested with Not I and cloned into
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."

ORIGIN

Query Match 35.3%; Score 850.4; DB 13; Length 918;
 Best Local Similarity 97.9%; Pred. No. 4.7e-217;
 Matches 868; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

QY 864 AAAGGCTGCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTGAAGCGGAAACAGCTTGA 923
 |||||
 Db 21 AAAGGCTGCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTGAAGCGGAAACAGCTTGA 80
 |||||
 QY 924 TGGAGAAAACCTCGTACAGTGCAGAAAGTGTAAAGATGGTTCACGTTCAAAAGAGGTT 983
 |||||
 Db 81 TGGAGAAAACCTCGTACAGTGCAGAAAGTGTAAAGATGGTTCACGTTCAAGAGGTT 140
 |||||
 QY 984 CACTATCCATAGATCTCTTAATGTTCTTACACATTTCTCTGAAACGTTTTCGAAATTTTAC 1043
 |||||
 Db 141 CACTATCCATAGATCTCTTAATGTTCTTACACATTTCTCTGAAACGTTTTCGAAATTTTAC 200
 |||||
 QY 1044 CGGTGAAAATTTGCTAAGGATGTGAATACCCCTGAGTATCTTGAATTCGGCCATATAT 1103
 |||||
 Db 201 CGGTGAAAATTTGCTAAGGATGTGAATACCCCTGAGTATCTTGAATTCGGCCATATAT 260
 |||||
 QY 1104 GTCTCAACCCAGGAGGAGGCAATTTGTCTACGTCTTGTATGTCAGTGTGTCACACTGG 1163
 |||||
 Db 261 GTCTCAACCCAGGAGGAGGCAATTTGTCTACGTCTTGTATGTCAGTGTGTCACACTGG 320
 |||||
 QY 1164 TTTTAAATTCGCATGCTGGCCATTACTTCTGCTACATAAAAGCTAGCAATGGCCTCTGGTA 1223
 |||||
 Db 321 TTTTAAATTCGCATGCTGGCCATTACTTCTGCTACATAAAAGCTAGCAATGGCCTCTGGTA 380
 |||||
 QY 1224 TCAATATGAATGACTCCATTGTATCTACCAAGTATAGATCGGTACTCAGCCCAACAGC 1283
 |||||
 Db 381 TCAATATGAATGACTCCATTGTATCTACCAAGTATAGATCGGTACTCAGCCCAACAGC 440
 |||||
 QY 1284 CTATGTGCTCTTTTATATACAGTCCCATGATGTGAAAAATGAGGTGAATTTACTCATCC 1343
 |||||
 Db 441 CTATGTGCTCTTTTATATACAGTCCCATGATGTGAAAAATGAGGTGAATTTACTCATCC 500
 |||||
 QY 1344 CACCCATAGCCCCGCCAGTCTCTCCCGCCCGCTCATCTAGTCAGCGGGTTGTCCACAA 1403
 |||||
 Db 501 CACCCATAGCCCCGCCAGTCTCTCCCGCCCGCTCATCTAGTCAGCGGGTTGTCCACAA 560
 |||||
 QY 1404 CAAACAGCTCGCCAGGCTTTTATCGGACCACAGCTTCCCTCTCATATGATAAAGAAATCC 1463
 |||||
 Db 561 CAAACAGCTCGCCAGGCTTTTATCGGACCACAGCTTCCCTCTCATATGATAAAGAAATCC 620
 |||||
 QY 1464 ACCTCACTTAATGGGACTGGACCAATTGAAAGACACGCCAAGCAGTTCATGTGAGTCC 1523
 |||||
 Db 621 ACCTCACTTAATGGGACTGGACCAATTGAAAGACACGCCAAGCAGTTCATGTGAGTCC 680
 |||||
 QY 1524 TAACGGGAATTCAGTGTCAACAGGCTAGTCTCTGTTAATGCTTCAGTCTCTGTCACAAA 1583
 |||||
 Db 681 TAACGGGAATTCAGTGTCAACAGGCTAGTCTCTGTTAATGCTTCAGTCTCTGTCACAAA 740
 |||||
 QY 1584 CTGTCAGTTAATAGGTCCTCAGTGATCCCAAGAACATCTTAAGAAACAAAAAATTACAAAT 1643
 |||||
 Db 741 CTGTCAGTTAATAGGTCCTCAGTGATCCCAAGAACATCTTAAGAAACAAAAAATTAC-AT 799
 |||||
 QY 1644 CAGTATTCAACAAAGTTGCTCTGTCGCGAGTGTCACTCAACCTAACCTTCATAGTAA 1703
 |||||
 Db 800 CAGTATTCAACAAAGTTGCTCTGTCGCGAGTGTCACTCAACCTAACCTTCATAGTAA 859
 |||||
 QY 1704 TTCTTTGGAGAACCTTACCAAGCCGCTTCCTCTTCTACCATTACCA 1750
 |||||
 Db 860 NTCTTTGGAGAACCTTACCAAGCCGCTTCCTCTTCTACCATTACCA 906
 |||||

RESULT 8

LOCUS BUI73371
 DEFINITION BUI73371.1 GI:22687355
 5', mRNA sequence.
 ACCESSION BUI73371
 VERSION BUI73371.1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 894)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM13548 row: g column: 07
 High quality sequence stop: 670.
 Location/Qualifiers
 1..894
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6174750"
 /tissue_type="leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_71"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 kb."

FEATURES
 source
 1..894
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6174750"
 /tissue_type="leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_71"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 kb."

ORIGIN
 Query Match 34.7%; Score 836.4; DB 13; Length 894;
 Best Local Similarity 98.9%; Pred. No. 2.7e-213; Indels 4; Gaps 4;
 Matches 884; Conservative 0; Mismatches 6; Indels 4; Gaps 4;
 109 CCATCAGCCTATCAGAAATCAGCTGGCAGCTCCGAGCAGTCTCAGCTGGAGCATGGAT 168
 1 CCATCAGCCTATCAGAAATCAGCTGGCAGCTCCGAGCAGTCTCAGCTGGAGCATGGAT 60
 169 GCAGGTTCTCCAGCTGGGCTGCTGCTTCATGTAATGATGTCGAATCAGCACTT 228
 61 GCAGGTTCTCCAGCTGGGCTGCTGCTTCATGTAATGATGTCGAATCAGCACTT 120
 229 TCTTTAGGACAGTACCTGGTGTGTTATTCGAGTTTCATCTGACCTGATAATCA 288
 121 TCTTTAGGACAGTACCTGGTGTGTTATTCGAGTTTCATCTGACCTGATAATCA 180
 289 AAACCATCACCACAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCTCCACAGAAAGTT 348
 181 AAACCATCACCACAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCTCCACAGAAAGTT 240
 349 CTTTTCCTCATCTGAGAAGATTGTTTAAAGTGGCAACAACTCATAGATTGGAGCTGGG 408
 241 CTTTTCCTCATCTGAGAAGATTGTTTAAAGTGGCAACAACTCATAGATTGGAGCTGGG 300
 409 CTCAGAAATTTGGCAATACCTGTTTCCCAATGACGACGTCAGTGTAACTACACACA 468
 301 CTCAGAAATTTGGCAATACCTGTTTCCCAATGACGACGTCAGTGTAACTACACACA 360
 469 CCACCTCTTGCCAAATTCATGCTATCATGATCACTCAAAACATGTCATGCGAAGGC 528
 361 CCACCTCTTGCCAAATTCATGCTATCATGATCACTCAAAACATGTCATGCGAAGGC 420
 529 TTTTGTATGATGTGACAAATGCAAGCACAATATACCGAGGCACTCAGTAATCTGGGAC 588
 421 TTTTGTATGATGTGACAAATGCAAGCACAATATACCGAGGCACTCAGTAATCTGGGAC 480
 589 GTTATTAACCAATGTTTGTCAATCAATGATGCGGCGGTATAGCTAGGCACTCCGTTT 648
 481 GTTATTAACCAATGTTTGTCAATCAATGATGCGGCGGTATAGCTAGGCACTCCGTTT 540
 649 GGAAACCAAGAGATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGCAAGACA 708
 541 GGAAACCAAGAGATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGCAAGACA 600

QY 709 TGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCACCACTCTTTGTTGTCAG 768
 Db 601 TGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCACCACTCTTTGTTGTCAG 660
 QY 769 ATATTGGAGGATACCTAAGATCTAGAGTCAAAATGTTTAAATTTGCAAGGGGTTTCAGAT 828
 Db 661 ATATTGGAGGATACCTAAGATCTAGAGTCAAAATGTTTAAATTTGCAAGGGGTTTCAGAT 720
 QY 829 ACTTTTGTATCCATATCTTTGATATACATTTGGAGATAAAGGGTGTCTCAGAGTGTCAACAG 888
 Db 721 ACTTTTGTATCCATATCTTTGATATACATTTGGAGATAAAGGGTGTCTCAGAGTGTCAACAG 780
 QY 889 GCATTGGAGCAG-TTTGTGAAGCCGGAACAGCTTGTATGG-AGAAAATCTGTAACAGTGA 946
 Db 781 GCATTGGAGCAGTTTGTGAAGCCGGAACAGCTTGTATGGNAGAAAATCTGTAACAGTGA 840
 QY 947 GCAAGT-GTAAAAAGATGTTCCAGC-TTCAAGAGGTTTCATCTCCATAGATC 998
 Db 841 GCAAGTGGTAAAAAGATGTTCCAGCTTTCCAAAGAGTTCATCTATTCATAGAAC 894

RESULT 9
 AU124746 851 bp mRNA linear EST 01-AUG-2002
 LOCUS AU124746 NT2RM4 Homo sapiens cDNA clone NT2RM4000488 5', mRNA
 DEFINITION sequence.
 ACCESSION AU124746
 VERSION AU124746
 KEYWORDS AU124746.1 GI:10949462
 SOURCE EST.
 ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 851)
 AUTHORS Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and Isogai,T.
 TITLE HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S., Isogai,T.)
 JOURNAL Unpublished (2000)
 COMMENT Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3975
 Fax: 81-438-52-3986
 Email: genomics@hri.co.jp
 HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES
 Location/Qualifiers
 1..851
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="NT2RM4000488"
 /cell_type="teratocarcinoma"
 /cell_line="NT2"
 /clone_lib="NT2RM4"
 /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal precursor cells"

ORIGIN
 Query Match 33.6%; Score 811.2; DB 9; Length 851;
 Best Local Similarity 98.5%; Pred. No. 1.6e-206;
 Matches 838; Conservative 0; Mismatches 11; Indels 2; Gaps 2;
 1339 CATCCACCCATAGCCCGGCGAGTCCTCTCCCGCCCGCTCATCAGTCAGGGGTGTC 1398
 2 CATCCACCCATAGCCCGGCGAGTCCTCTCCCGCCCGCTCATCAGTCAGGGGTGTC 61

Wed Aug 18 13:53:02 2004

QY	1935	GCTCCAGCCGCTGATGAATGGCAATCAAGCTGAACTCCAGCGTGTGGT-GCCCTAT	1893	Db	121	CATGATCAGGTTCTGCGAGCTGGGGTGGCTGTCTTCATGAATGATGTCAATCA	180
Db	717	GCKCCAGCCGCTGATGAATGGCAATCAAGCTGAACTCCAGCGTGTGGTGGCCCTAT	776	QY	222	CACACTTTCTTTAGACAGTACCTGGTGGTGTAGTTATTCGAGTTTCATCTGTACCTGA	281
QY	1894	GGCGCCGAGTCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGCAAGAGAAAT-GG	1952	Db	181	CACACTTTCTTTAGACAGTACCTGGTGGTGTAGTTATTCGAGTTTCATCTGTACCTGA	240
Db	777	GGCGCCGAGTCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGCAAGAGAAATGGG	836	QY	282	TAAATCAAAACCATCACCAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCTCCACA	341
QY	1953	GATTGATGAGTGTGAGTCTCCACTCTCCCGCCCAAGATGCCAGATGAGGAGGCCAC	2012	Db	241	TAAATCAAAACCATCACCAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCTCCACA	300
Db	837	GATTGATGAGTGTGAGTCTCCACTCTCCCGCCCAAGATGCCAGATGAGGAGGCCA-	895	QY	342	GAAAGTCTTTTCCCATCTGAGAGATTTTGTCTTAAGTGGCAACAACTCATAGAGTTGG	401
QY	2013	TCCGACGAGTCTCAAGAACCATGACCTCAACCGTGTCTAATAGT	2058	Db	301	GAAAGTCTTTTCCCATCTGAGAGATTTTGTCTTAAGTGGCAACAACTCATAGAGTTGG	360
Db	896	TDCGACGAGTCTCAAGAACCATGACCTCAACCGTGTCTAATAGT	941	QY	402	AGCTGGGCTCCAGATTTTGGCAATACCTGTTTGGCAATGCGACACTGCAGTGTTAAC	461
RESULT 11				Db	361	AGCTGGGCTCCAGATTTTGGCAATACCTGTTTGGCAATGCGACACTGCAGTGTTAAC	420
LOCUS	BUI195574	880 bp	mRNA	QY	462	CTACACACACCTCTTGGCAATTTACATGCTATCAATGAGATGCGGGCTATAGCTAGGCCT	521
DEFINITION	AGENCOURT_7967650 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6171493		EST 04-SEP-2002	Db	421	CTACACACACCTCTTGGCAATTTACATGCTATCAATGAGATGCGGGCTATAGCTAGGCCT	480
ACCESSION	BUI195574			QY	522	AGAAGGCTTTTGTATGATGTGTAATGCAAGCACAATATTACCCAGGCACTCAGTAATCC	581
VERSION	BUI195574.1	GI:22709558		Db	481	AGAAGGCTTTTGTATGATGTGTAATGCAAGCACAATATTACCCAGGCACTCAGTAATCC	540
KEYWORDS	EST.			QY	582	TGGGGAGCTTTATTAACCAATGTTTGTTCATCAATGAGATGCGGGCTATAGCTAGGCCT	641
SOURCE	Homo sapiens (human)			Db	541	TGGGGAGCTTTATTAACCAATGTTTGTTCATCAATGAGATGCGGGCTATAGCTAGGCCT	600
ORGANISM	Homo sapiens			QY	642	CCGTTTGGAAACCAAGAGATGCCAATGAAATTCCTTCAATACACTCTTGTATGCTATGCA	701
REFERENCE	NIH-MGC http://mgi.nci.nih.gov/.			Db	601	CCGTTTGGAAACCAAGAGATGCCAATGAAATTCCTTCAATACACTCTTGTATGCTATGCA	660
AUTHORS	1 (bases 1 to 880)			QY	702	GAAAGCATGCTTGAATGGGAGCAATAATTAGACAGACACACCCAGGCACTCTTGT	761
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			Db	661	GAAAGCATGCTTGAATGGGAGCAATAATTAGACAGACACACCCAGGCACTCTTGT	720
COMMENT	Unpublished (1999)			QY	762	TTGTCAGATATTTTGAGGATACCTTAAGATCTAGAGTCAAAATGTTTAAATTCGAGGGCGT	821
	Contact: Robert Strausberg, Ph.D.			Db	721	TTGTCAGATATTTTGAGGATACCTTAAGATCTAGAGTCAAAATGTTTAAATTCGAGGGCGT	780
	Email: cgabs-remail.nih.gov			QY	822	TTGTCAGATATTTTGAGGATACCTTAAGATCTAGAGTCAAAATGTTTAAATTCGAGGGCGT	859
	Tissue Procurement: ATCC			Db	781	TTGTCAGATATTTTGAGGATACCTTAAGATCTAGAGTCAAAATGTTTAAATTCGAGGGCGT	
	cDNA Library Preparation: Life Technologies, Inc.			RESULT 12			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			AUI124710			
	DNA Sequencing by: Agencourt Bioscience Corporation			LOCUS	AUI124710	822 bp	linear
	Clone distribution: MGC clone distribution information can be			DEFINITION	AUI124710 NT2RM4 Homo sapiens cDNA clone NT2RM4000431 5', mRNA		EST 01-AUG-2002
	found through the I.M.A.G.E. Consortium/LLNL at:			ACCESSION	AUI124710		
	http://image.llnl.gov			VERSION	AUI124710.1	GI:10949426	
	Plate: LLNL3539 row: 0 column: 14			KEYWORDS	Homo sapiens (human)		
	High quality sequence stop: 710.			SOURCE	Homo sapiens		
FEATURES	Location/Qualifiers			ORGANISM	Homo sapiens		
source	1..880			REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	/organism="Homo sapiens"			AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	/mol_type="mRNA"			TITLE	Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J.,		
	/db_xref="taxon:9606"			JOURNAL	Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S. and		
	/clone="IMAGE:6171493"			COMMENT	Isogai, T.		
	/tissue_type="retinoblastoma"				HPI human cDNA project (Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S.,		
	/lab_host="DH10B (phage-resistant)"				Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T.,		
	/clone_lib="NIH_MGC_67"				Suzuki, Y., Sugano, S., Isogai, T.)		
	/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;				Unpublished (2000)		
	Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.				Contact: Takao Isogai		
	Average insert size 1.75 kb. Library constructed by Life				Genomics Laboratory		
	Technologies."				Helix Research Institute		
					1532-3 Yana, Kisarazu, Chiba 292-0812, Japan		
					Tel: 81-438-52-3975		
ORIGIN							
Query Match	33.3%;	Score	803.8;	DB 13;	Length	880;	
Best Local Similarity	99.6%;	Pred. No.	1.6e-204;	Indels	1;	Gaps	1;
Matches	816;	Conservative	0;	Mismatches	2;		
QY	42	ACGAGCGCCGAGCGGTCAGAGTTGAAACATGACCATGATGTTGACAAAGCTTCTGAATC	101	Db	42	ACGAGCGCCGAGCGGTCAGAGTTGAAACATGACCATGATGTTGACAAAGCTTCTGAATC	60
QY	102	TTCAGACCCATCAGCCATCAGAAATCAGCTGCGAGTCCGAGGCGATCTCAGCTGAGA	161	QY	102	TTCAGACCCATCAGCCATCAGAAATCAGCTGCGAGTCCGAGGCGATCTCAGCTGAGA	161
Db	61	TTCAGACCCATCAGCCATCAGAAATCAGCTGCGAGTCCGAGGCGATCTCAGCTGAGA	120	Db	61	TTCAGACCCATCAGCCATCAGAAATCAGCTGCGAGTCCGAGGCGATCTCAGCTGAGA	120
QY	162	CATGGATGACAGGTTCTGCCAGCTGGGGTGTGTCTTCATTGAATGATGTGCAATCA	221	QY	162	CATGGATGACAGGTTCTGCCAGCTGGGGTGTGTCTTCATTGAATGATGTGCAATCA	221

Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1. .822
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RM4000431"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/clone_lib="NT2RM4"
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
precursor cells"

ORIGIN

Query Match 33.0%; Score 795.8; DB 9; Length 822;
Best Local Similarity 98.8%; Pred. No. 2.2e-202;
Matches 811; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 856 TTGGAGATTAAGGCTGCTCAGAGTGTCACAAAGGCAATGGAGCAGTTTGTGAAGCCGGA 915
Db |||||
QY 916 CAGCTTTGATGGAGAAACCTGCTACAGTGCACCAAGTGTAAAGAGTGTTCAGCTTCA 975
Db |||||
QY 61 CAGCTTTGATGGAGAAACCTGCTACAGTGCACCAAGTGTAAAGAGTGTTCAGCTTCA 120
Db |||||
QY 976 AAGAGTTCACTATCCATAGATCCCTTAATGTTCTTACACTTTCTGAAAGCTTTGCA 1035
Db |||||
QY 121 AAGAGTTCACTATCCATAGATCCCTTAATGTTCTTACACTTTCTGAAAGCTTTGCA 180
Db |||||
QY 1036 AATTTACGGTGGAAAAATTGCTAAGAGTGTGAATACCCCTGAGTATCTTGATATCGG 1095
Db |||||
QY 181 AATTTACGGTGGAAAAATTGCTAAGAGTGTGAATACCCCTGAGTATCTTGATATCGG 240
Db |||||
QY 1096 CCATATATGTTCTCAACCAAGGAGAGCAATGTCTAGCTTTGTATGCGTGTGTC 1155
Db |||||
QY 241 CCATATATGTTCTCAACCAAGGAGAGCAATGTCTAGCTTTGTATGCGTGTGTC 300
Db |||||
QY 1156 CACACTGGTTTAAATGTCATGCTGGCCATTAATGTTCTGCTACATAAAGCTTAGCAATGGC 360
Db |||||
QY 301 CACACTGGTTTAAATGTCATGCTGGCCATTAATGTTCTGCTACATAAAGCTTAGCAATGGC 360
Db |||||
QY 1216 CTCTGGTATCAATGAATGATCCATGTTATCTACAGTGTATATAGATCGGTACTCAGC 1275
Db |||||
QY 361 CTCTGGTATCAATGAATGATCCATGTTATCTACAGTGTATATAGATCGGTACTCAGC 420
Db |||||
QY 1276 CAACAAGCCTATGCTCTTTTATATCAGGTCCCATGATGTGAATAATGGAGGTGAACCT 1335
Db |||||
QY 421 CAACAAGCCTATGCTCTTTTATATCAGGTCCCATGATGTGAATAATGGAGGTGAACCT 480
Db |||||
QY 1336 ACTCATCCCAACCCATAGCCCGGCGAGTCTCTCCCGCCCGTCATCAGTCAGCGGGTT 1395
Db |||||
QY 481 ACTCATCCCAACCCATAGCCCGGCGAGTCTCTCCCGCCCGTCATCAGTCAGCGGGTT 540
Db |||||
QY 1396 GTCAACCAACAAACAGGCTGGCGCCAGGCTTTATCGGACCAAGCTTCCCTCTCATCATGA 1455
Db |||||
QY 541 GTCAACCAACAAACAGGCTGGCGCCAGGCTTTATCGGACCAAGCTTCCCTCTCATCATGA 600
Db |||||
QY 1456 AAGAATCCACTCACTTAAATGGAGTGGACCATTTGAAGACACAGCCAGGTTCCATG 1515
Db |||||
QY 601 AAGAATCCACTCACTTAAATGGAGTGGACCATTTGAAGACACAGCCAGGTTCCATG 660
Db |||||
QY 1516 TCAGTCTTAACGGGAATCCAGTGTCAACAGGGCTAGTCTCTGTTAATGCTTCAGTTCT 1575
Db |||||
QY 661 TCAGTCTTAACGGGAATCCAGTGTCAACAGGGCTAGTCTCTGTTAATGCTTCAGTTCT 720
Db |||||
QY 1576 GTCCAAACCTGGTCACTTAATAGGTCTCTAGTATCCGAAACATCTTAAGAAACAAAA 1635
Db |||||

Db 721 GTCAAAACCTGGGCGAGTTAATAGTCTCTCA-TGATCCAGAACATCTTAAGAAACCAAA 779
QY 1636 ATTACATCAGTATTATCAACAAGTTCCTGTTCCGCCAGTG 1676
Db |||||
780 ATTCAATCAGTATTATCAACAAGTTCCTGTTCCGCCANGG 820

RESULT 13
AUI21934
LOCUS AUI21934
DEFINITION MAMMAL Homo sapiens cDNA clone MAMMAL001308 5', mRNA
sequence.
AUI21934 818 bp mRNA linear EST 01-AUG-2002
ACCESSION AUI21934
VERSION AUI21934.1 GI:10937169
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 818)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1. .818
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MAMMAL001308"
/tissue_type="mammary gland"
/clone_lib="MAMMAL"
/note="Vector: pME18SFL3"

ORIGIN

Query Match 31.6%; Score 762; DB 9; Length 818;
Best Local Similarity 98.5%; Pred. No. 2.7e-193;
Matches 809; Conservative 0; Mismatches 8; Indels 4; Gaps 4;

QY 727 AAATTAGACAGACACACCCAGGCCACCACTCTTGTGTGTCAGATATTGGAGGATACCTA 786
Db 1 AAATTAGACAGACACACCCAGGCCACCACTCTTGTGTGTCAGATATTGGAGGATACCTA 60
QY 787 AGATCTAGAGTCAAAATGTTTAAATTGCAAGGGCTTTTCAGATCTTTTTCATCATCTT 846
Db 61 AGATCTAGAGTCAAAATGTTTAAATTGCAAGGGCTTTTCAGATCTTTTTCATCATCTT 120
QY 847 GATATAACATTTGGAGTAAAGGCTGTCTCAGAGTGTCAACAGGCAATTTGGAGCTTTGTG 906
Db 121 GATATAACATTTGGAGTAAAGGCTGTCTCAGAGTGTCAACAGGCAATTTGGAGCTTTGTG 180
QY 907 AAGCCGAAACAGCTTGTATGGAGAAACTGTCTAAGTGTGAGAGTGTAAAAAGATGGTT 966
Db 181 AAGCCGAAACAGCTTGTATGGAGAAACTGTCTAAGTGTGAGAGTGTAAAAAGATGGTT 239
QY 967 CCAGCTTTCAAAGAGGTTCATCTCCATAGATCTCTTAATGTTCTTACACTTCTCTGAAA 1026
Db 240 CCAGCTTTCAAAGAGGTTCATCTCCATAGATCTCTTAATGTTCTTACACTTCTCTGAAA 299
QY 1027 CGTTTTCGAAATTTTACCAGGTTGAAAAATTTGTAAGGATGTGAAATACCTGAGTATCTT 1086
Db |||||

/note="vector: pME18SFL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"

ORIGIN	Query Match	Score	DB	Length
	Best Local Similarity	30.4%	732.2	9
	Matches	745	Conservative	0
		Mismatches	6	
		Indels	1	
		Gaps	1	
QY	1166	TTAATGGCCATGCTGGCCATTTACTTCTGTACATATAAGCTAGCAATGSCCTCTCTGGTATC	1225	
DB	1	TTAATGGCCATGCTGGCCATTTACTTCTGTACATATAAGCTAGCAATGSCCTCTCTGGTATC	60	
QY	1226	AAATGAATGACTCCCATTTGTATCTACAGTGATTTAGATCGTACTCAGCCCAACAGCCT	1285	
DB	61	AAATGAATGACTCCCATTTGTATCTACAGTGATTTAGATCGTACTCAGCCCAACAGCCT	120	
QY	1286	ATGTGCTCTTTTATATATCAGTCCCATGATGTGAAATGAGAGTGAATCTTACTCATCCCA	1345	
DB	121	ATGTGCTCTTTTATATATCAGTCCCATGATGTGAAATGAGAGTGAATCTTACTCATCCCA	180	
QY	1346	CCCATAGCCCCCGCCAGTCTCTCTCCCGCCCGCTCATCATAGTCAGCGGGTTGTCAACACA	1405	
DB	181	CCCATAGCCCCCGCCAGTCTCTCTCCCGCCCGCTCATCATAGTCAGCGGGTTGTCAACACA	240	
QY	1406	AAAGCTGCGCCAGGCTTTATCGGACCAAGCTTCCCTCTCATCATGATAAAGATCCAC	1465	
DB	241	AAAGCTGCGCCAGGCTTTATCGGACCAAGCTTCCCTCTCATCATGATAAAGATCCAC	300	
QY	1466	CTCAGTTAAATGGAGTGGACCAATTTGAAAGACAGCCAGCAGTTCATGTCGAGTCTTA	1525	
DB	301	CTCAGTTAAATGGAGTGGACCAATTTGAAAGACAGCCAGCAGTTCATGTCGAGTCTTA	360	
QY	1526	ACGGGAATTCAGTGTCAACAGGCGCTAGTCTCTGTTAATGTTTCTGCTTCTGTCACAACT	1585	
DB	361	ACGGGAATTCAGTGTCAACAGGCGCTAGTCTCTGTTAATGTTTCTGCTTCTGTCACAACT	420	
QY	1586	GGTCAGTTAAATGAGTCTCTAGTATCCAGAACATCTCTAGAACACAAATTAACAATCA	1645	
DB	421	GGTCAGTTAAATGAGTCTCTAGTATCCAGAACATCTCTAGAACACAAATTAACAATCA	480	
QY	1646	GTATTCAACAAAGTGGCTGTTGCGCCAGTGTGAGTCTCAACCTTAACTTCTCATAGTAAT	1705	
DB	481	GTATTCAACAAAGTGGCTGTTGCGCCAGTGTGAGTCTCAACCTTAACTTCTCATAGTAAT	540	
QY	1706	CTTTGGAGAACCCCTACCAAGCCGCTTCTCTTCTTACCATTAACATTCGAGTACAGT	1765	
DB	541	CTTTGGAGAACCCCTACCAAGCCGCTTCTCTTCTTACCATTAACATTCGAGTACAGT	600	
QY	1766	CTACTCGAAGCCATCTACAGTGTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT	1825	
DB	601	CTACTCGAAGCCATCTACAGTGTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT	660	
QY	1826	GTGAATCTGCTCTCCAGCCGCTGATGAATGCAATCAAGCTCAAGCTCCAGCTGCTGG	1885	
DB	661	GTGAATCTGCTCTCCAGCCGCTGATGAATGCAATCAAGCTCAAGCTGCTGACTCA	719	
QY	1886	TGCCCTATGGCCGCGAGTCTCTCTGAGGACTCT	1917	
DB	720	TGCCCTATGGCCGCGAGTCTCTCTGAGGACTCT	751	

RESULT 15
 BM563326
 LOCUS
 DEFINITION
 AGENCOURT 6589448 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5441601
 5', mRNA sequence.
 ACCESSION
 VERSION
 BM563326
 BM563326.1
 GI:18810154
 EST.
 KEYWORDS
 Homo sapiens (human)
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Db	300	CGTTTTCGCAATTTTACCGTGTGAAAAATGCTAAGGATGGAATACCCCTGAGTATCTT	359
QY	1087	GATATTCGGCCATATATGCTCAACCCCAAGAGCCAAATTTGTCTAGCTCTTGATGCA	1146
DB	360	GATATTCGGCCATATATGCTCAACCCCAAGAGCCAAATTTGTCTAGCTCTTGATGCA	419
QY	1147	GTGCTGGTCCACACTGGTTTTAAATGGCCATGCTGCGCCATTTACTTCTGCTACATAAAGCT	1206
DB	420	GTGCTGGTCCACACTGGTTTTAAATGGCCATGCTGCGCCATTTACTTCTGCTACATAAAGCT	479
QY	1207	AGCAATGGCCCTGCTGATCAAAATGATGACTCCATTTGTATCTACAGTGATTTAGATCG	1266
DB	480	AGCAATGGCCCTGCTGATCAAAATGATGACTCCATTTGTATCTACAGTGATTTAGATCG	539
QY	1267	GTACTCAGCCACACAGCTATGCTCTTTTATATCAGTCCCATGATGTGAAAAATGGA	1326
DB	540	GTACTCAGCCACACAGCTATGCTCTTTTATATCAGTCCCATGATGTGAAAAATGGA	599
QY	1327	GGTGAATTTACTCATCCCAACCCATAGCCCGCCAGTCTCTCCCGCCCGCTCATCAGT	1386
DB	600	GGTGAATTTACTCATCCCAACCCATAGCCCGCCAGTCTCTCCCG-CCCCGTCATCAGT	658
QY	1387	CAGCGGTTGTACCAACCAAGCTGCTGCTTTTATATCAGTCCCATGATGTGAAAAATGGA	1446
DB	659	CAGCGGTTGTACCAACCAAGCTGCTGCTTTTATATCAGTCCCATGATGTGAAAAATGGA	718
QY	1447	CACATGATAAAGATCCACCTCACTTAAATGGAGTGGACCTTGAACATGGAACAGCCAGC	1506
DB	719	TACATGATAAAGATCC-CTTCACTTAAATGGAGTGGACCTTGAACATGGAACAGCCAGC	777
QY	1507	AGTTCCATGTCAGTCTTAAC-GGGAATTCAGTGTCAACA	1546
DB	778	AGTTNCATGTTNAGTCTTAACGGGAATTTTCAAGTGTCAACA	818

RESULT 14
 AUI30866
 LOCUS
 DEFINITION
 AUI30866 NT2RP3 Homo sapiens cDNA clone NT2RP3001577 5', mRNA
 sequence.
 AUI30866
 AUI30866.1
 GI:10991220
 EST.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 763)
 Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,
 Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and
 Isogai, T.
 HRI human cDNA project
 Unpublished (2000)
 CONTACT
 Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3975
 Fax: 81-438-52-3986
 Email: genomics@hri.co.jp
 HRI human cDNA project; 5'- and 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.
 FEATURES
 Location/Qualifiers
 1..763
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="NT2RP3001577"
 /cell_type="teratocarcinoma"
 /cell_line="NT2"
 /clone_lib="NT2RP3"

REFERENCE 1 (bases 1 to 1088)
NIH-MGC <http://mhc.nci.nih.gov/>;
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: ggapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
cDNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LLCM1916 row: c column: 10
High quality sequence stop: 611.
FEATURES
Location/Qualifiers
1..1088
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5441601"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 98"
/note="Organ: Brain; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Query Match 29.4%; Score 709.6; DB 12; Length 1088;
Best Local Similarity 99.3%; Pred. No. 4e-179;
Matches 723; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
QY 133 GGCAGCTCCGAGGCGAGTCTCAGCTGGAGACATGGATGAGGTTCTGCCAGCTGGGGTGCT 192
DB 1 GGCAGCTCCGAGGCGAGTCTCAGCTGGAGACATGGATGAGGTTCTGCCAGCTGGGGTGCT 60
QY 193 GTGCTTCATTGAATGATGTGTCAAAATCACACACTTTCTTTAGGACCAGTACCTGGTGCT 252
DB 61 GTGCTTCATTGAATGATGTGTCAAAATCACACACTTTCTTTAGGACCAGTACCTGGTGCT 120
QY 253 GTAGTTTATTCGAGTTCTCTGTACTGATTAATCAAAACCATCACCAAAAGGATCAA 312
DB 121 GTAGTTTATTCGAGTTCTCTGTACTGATTAATCAAAACCATCACCAAAAGGATCAA 180
QY 313 GCCCTAGGTGATGGCATCGCTCCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTGT 372
DB 181 GCCCTAGGTGATGGCATCGCTCCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTGT 240
QY 373 CTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGT 432
DB 241 CTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGT 300
QY 433 TTTCGCAATGCAGCAGCTGAGGTTTAACCTTACACACACCTCTTTGCCAATTAATGCTA 492
DB 301 TTTCGCAATGCAGCAGCTGAGGTTTAACCTTACACACACCTCTTTGCCAATTAATGCTA 360
QY 493 TCACATGAACACTCCAAACATGTCTGACAGAGGCTTTTGTATGATGTGTACAAATGCAA 552
DB 361 TCACATGAACACTCCAAACATGTCTGACAGAGGCTTTTGTATGATGTGTACAAATGCAA 420
QY 553 GCACATATTACCCAGGCACTCAGTAATCTTGGGACGTTATTAAACCAATGTTTGTGATC 612
DB 421 GCACATATTACCCAGGCACTCAGTAATCTTGGGACGTTATTAAACCAATGTTTGTGATC 480
QY 613 AATGAGATGGCGGTATAGCTAGGCACTCGCTTTTGGAAACCAAGAGATGCCCATGAA 672
DB 481 AATGAGATGGCGGTATAGCTAGGCACTCGCTTTTGGAAACCAAGAGATGCCCATGAA 540

QY 673 TTCCTTCAATACACTGTTGATGCTATGCAGAAAGCATGCTTGAATGGCAGCAATAAATTA 732
DB 541 TTCCTTCAATACACTGTTGATGCTATGCAGAAAGCATGCTTGAATGGCAGCAATAAATTA 600
QY 733 GACAGACACACCCAGGCCACACACTCTTTGTTGTCAGATATTTGGAGGATACCTAAGATCT 792
DB 601 GACAGACACACCCAGGCCACACACTCTTTGTTGTCAGATATTTGGAGGATACCTAAGATCT 660
QY 793 AGAGTCAAAATGTTTAA-TTTGCAGGGCGTTTTCAGATACCTTTGATCCATATCTTGATAT 851
DB 661 AGAGTCAAAATGTTTAAATTTGCAAGGCGGTTTCAATACTTTTGCATATCTTGATAT 720
QY 852 AACATTGG 859
DB 721 AACATTGG 728

Search completed: August 16, 2004, 20:20:44
Job time : 9145 secs

THIS PAGE BLANK (USPTO)

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	474.6	19.7	722	4	US-09-280-116-223		Sequence 223, App
2	421.8	17.5	3941	4	US-09-408-865-2		Sequence 2, Appli
3	393.8	16.3	457	4	US-09-280-116-214		Sequence 214, App
4	320	13.3	917	4	US-09-280-116-319		Sequence 219, App
5	313.6	13.0	1873	3	US-09-019-095A-37		Sequence 37, Appli
6	300	12.4	2674	3	US-09-019-095A-1		Sequence 1, Appli
7	283.6	11.8	1523	3	US-09-019-095A-21		Sequence 21, Appl
8	194.2	8.1	672	4	US-09-280-116-217		Sequence 217, App
9	47.2	2.0	8096	4	US-09-058-489-33		Sequence 33, Appl
10	44.8	1.9	10091	3	US-09-058-489-34		Sequence 34, Appl
11	42.2	1.7	1998	4	US-09-280-116-222		Sequence 222, App
12	42.2	1.7	5734	4	US-09-976-594-862		Sequence 862, App
13	41.8	1.7	1683	4	US-09-976-594-861		Sequence 861, App
C 14	38.2	1.6	730	4	US-08-858-207A-234		Sequence 234, App
C 15	38	1.6	847	4	US-09-495-050A-40		Sequence 40, Appl
16	37.6	1.6	1074	4	US-09-489-039A-6272		Sequence 6272, Ap
17	37.4	1.6	1273	4	US-09-976-594-448		Sequence 448, App
18	37	1.5	283	4	US-09-313-294A-2500		Sequence 2500, App
19	36.8	1.5	3660	4	US-09-107-532A-2366		Sequence 2366, Ap
20	36.6	1.5	9834	4	US-08-956-171E-37		Sequence 37, Appl
C 21	36	1.5	4980	4	US-09-134-000C-2969		Sequence 2969, Ap
C 22	35.6	1.5	2197	4	US-09-778-171-1		Sequence 1, Appli
C 23	35.4	1.5	399	4	US-09-621-976-8976		Sequence 8976, Ap
24	35	1.5	19446	4	US-08-961-527-51		Sequence 51, Appl
25	34.8	1.4	832	4	US-09-621-976-2813		Sequence 2813, Ap
C 26	34.8	1.4	4029	4	US-09-620-312D-201		Sequence 201, App
C 27	34.8	1.4	28720	4	US-09-341-587-7		Sequence 7, Appli

407 AGTCCTCTCCCGCCCGCTCAGTCACGCGGTTGTACCAACAAACAGGCTGCG-CAG 465
 1421 GCTTTATCGGACACAGCTTCCCTCTCATATGATAAAGAAATCCACTCACTTAATGGGA 1480
 466 GCTTTATCGGACACAGCTTCCCTCTCATATGATAAAGAAATCCACTCACTTAATGGGA 525
 1481 C-TGGACCAATTCGAAGACACAGGCAAGCAGTTCATGTCAGTCCCTTAACGGGAATCCAGT 1539
 526 CATGACCAATTCGAAGACACAGGCAAGCAGTTCATGTCAGTCCCTTAACGGGAATCCAGT 585
 1540 GTCAACAGGCGTAGCTCTGTTAATGCTTCTGTCCTCAAACTGGTCAGTTAATAGG 1599
 586 GTCAACAGGCGTAG-CCTGTTAATGTT-CACTNWTGCCAAATGGTCAGTTAATAGG 640
 1600 TCTCTAGTCATCCAGAACATCTTAAGAAACAAATAATACATCATGATTTCAACAAG 1659
 641 T-CTCAGTGATCC--AGAATCTTANGAAACAAATAATCAANAGTATNACAACAAGTTG 697
 1660 TTGCC 1664
 698 CTGTC 702

RESULT 2

US-09-408-865-2
 ; Sequence 2, Application US/09408865A
 ; Patent No. 6329171
 ; GENERAL INFORMATION:
 ; APPLICANT: Kapeller-Libermann, Rosana
 ; TITLE OF INVENTION: 23484, A NOVEL HUMAN UBIQUITIN
 ; FILE REFERENCE: 5800-51
 ; CURRENT APPLICATION NUMBER: US/09/408.865A
 ; CURRENT FILING DATE: 1999-09-30
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 3941
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; NAME/KEY: CDS
 ; LOCATION: (279)...(3650)
 ; US-09-408-865-2

Query Match 17.5%; Score 421.8; DB 4; Length 3941;
 Best Local Similarity 63.8%; Pred. No. 7.8e-123;
 Matches 639; Conservative 0; Mismatches 362; Indels 0; Gaps 0;
 301 CAAAGGATCAAGCCCTAGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCT 360
 540 CACAGTATGAGAGCTGTGTGACGAGTCCAGCCCGCCAGAAAGTGTCTTTCCCGCAG 599
 361 GAGAGATTTGTTTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTTG 420
 600 GAGCGACTCTCTGAGGTGGAGCGGGTCTCCGCTGGGCGCAGGACTCCACACCTT 659
 421 GCAATACCTGTTTCCCAATGACGACTGAGTGTAACTTAACTTACACACACCTTTGCC 480
 660 GGCAACACCTTGCTTCTCAATGACCACTTCCAGTGTGTGACCTTACACACACCTTAGC 719
 481 AATTACATGCTATCATGAAACACTCCAAACATGTATGACAGAGGCTTTGTATGATG 540
 720 AACTACCTGCTCTCCAGGAGCATGTCTGAGCTGCCACAGGAAAGCTTCTCATGCTG 779
 541 TGTACATCAAGACACATATTTACCGGCACTCAGTAAATCTCGGGGACGTTTATTAACCA 600
 780 TGTGTCATGAGAACCACTTGTCCAGCTTCCGCAACAGCGGCAAGCCATCAAGCCC 839
 601 ATGTTTGTCAATAGATGACGCGGTATAGTACGACCTCGGTTTGGAAACCAAGAA 660
 840 GTCTCTTCTATCCGAGACCTGMAAAGATGCGCCGACACTTCCGCTTTGGGAACCGAG 899
 661 GATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGCAGAAAGCATGCTTGAATGGC 720

900 GACGCGCATGAGTCTCTGCGGTACACCATCGACGCAATGCAGAAAGCCCTGCTGATGCG 959
 721 AGCAATAATPAGACAGACACACCCAGGCGCCACTCTTTGTTGTGAGATATTTGGAGGA 780
 960 TGTGCAAGTTGGATGCTGTCAAACGAGGCTACTCTTGGTCCATCAAAATTTTGGAGGG 1019
 781 TACCTAAGATCTPAGAGTCAAAATGTTTAAATTCGAAGGGGCTTTCAGATACCTTTCATCCA 840
 1020 TATCTCAGATCACGCGTGAAGTGTCCGTTGTGCAAGAGCGTCTCGGACACCTACGACCCC 1079
 841 TATCTTGAATPAACTTGGAGATAAGGCTGTCTCAGAGTGTCAAAGCAATTTGGAGCAG 900
 1080 TACTTGGAGCTGCGCTGGAGATCCGGCAAGCTGCGAATATTGTGCGTCTCTGTHACTT 1139
 901 TTTGTCAAGCGGACAGCTTGTATGGAGAAATCTGTACAGTGCAGAGCAATTTGGAGCAG 960
 1140 TTTGTGAAAGCAGATGTCTCTGAGTGAGAGATGCCCTACATGTGTGCTAAATGCAAGAAG 1199
 961 ATGTTTCCAGCTTCAAAGAGGTTCACTATCCATAGATCTCTTAATGTTCTTACACTTTCT 1020
 1200 AAGTTTCCAGCGCAAGCGCTTCAACATCCACAGAAATCCAAACGCTTAAACCTTTCC 1259
 1021 CTGAACGTTTTCGAATTTTACCGGTGGAAATTTCTAAGGATGTGAATAATCCCTGAG 1080
 1260 CTCGAAGCGCTTTCGAATTTTACCGGTGGAAATTTCTAAGGATGTGAATAATCCCTGAG 1319
 1081 TATCTTGAATPAACTTGGAGATAAGGCTGTCTCAGAGTGTCAAAGCAATTTGGAGCAG 1140
 1320 TTTCTCAACATACATCTGCTGATATGTCCTCAGAAATTAATGCTGATCTCTGATATGAGCTC 1379
 1141 TATGAGTGTGTCACACCTGTTTAAATGCTGATGTCGCAATTTCTCTGCTTACATA 1200
 1380 TATGCTGTCTGTCGCTCTGAGTGTGAGAGATGATGATGATGATGATGATGATGATGATG 1439
 1201 AAAGCTAGCATGCTCTGATCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
 1440 AAGGCAAGCAATGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1499
 1261 AGATCGGTACTCAGCAACAGCCTATGCTGCTCTTTTATAT 1301
 1500 AAGTGGTCTTGACACAGCAGCGCTACGCTGCTTCTATCT 1540

RESULT 3

US-09-280-116-214
 ; Sequence 214, Application US/09280116A
 ; Patent No. 6331427
 ; GENERAL INFORMATION:
 ; APPLICANT: Robison, Keith E.
 ; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
 ; FILE REFERENCE: 5800-24, 035800/176965
 ; CURRENT APPLICATION NUMBER: US/09/280,116A
 ; CURRENT FILING DATE: 1999-03-26
 ; NUMBER OF SEQ ID NOS: 268
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 214
 ; LENGTH: 457
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: ubiquitin carboxyl-terminal hydrolases family 2
 ; US-09-280-116-214

Query Match 16.3%; Score 393.8; DB 4; Length 457;
 Best Local Similarity 37.2%; Pred. No. 1.3e-114;
 Matches 444; Conservative 0; Mismatches 7; Indels 6; Gaps 4;
 38 ATGAGCGAGCGCCGAGCGCGGTGAGATTTGAACATGACCATGATTTGACAAAGCTTCTG 97
 1 ATGAGCGAGCGCGAGCGCGGTGAGATTTGAACATGACCATGATTTGACAAAGCTTCTG 60
 98 AATCTTCAGACCCATCAGCCT--ATCAGAAATCAGCCTGGCAGCTCCGAGGCGAGTCTCACC 155

QY	892	TTGGAGCGATTGTGAAACCGGAAACAGCTTGTATGAGAGAAAAC	TGTAACAAGTCGACGAAG	951
Db	811	TTTGTGGGTACAGAGAAAGTCAGAACAGCTATGTGGAGATAAT	GCTACTACTCTGTGGTAAG	870
QY	952	TGTAAAAAGATGGTTCACGCTTCANAAGAGGTTCACTATCCATAGATCC	TCTTAATGTTCTTT	1011
Db	871	TGTAGACAGAAGATGCCAGCTTCTAAGACCCGTGCATGTCTCATAT	TGTGTCCTCAAGGTA	930
QY	1012	ACACTTTCTCTGAAACGTTTGTGCAAAATTTACCGGTGGAAAAAT	TGCTAAGGATGCGAA	1071
Db	931	ATGGTAGTGTAAATCGCTTCTCAGCCTTCACGGGTACAAAGT	TAGACAGAAAAGTAA	990
QY	1072	TACCTTGAGTATCTTGATATATTCGGCCATATATGTCTCAACCCAA	CGGAGAGCCAAATGTG	1131
Db	991	TACCGGAGTTCCTGTGACCTGAAGCCATACCTGTCTGAGCCTA	CTGTGAGGACCTTGCCT	1050
QY	1132	TACGTCCTTGATGACAGTGCTCGTCCACACTGGTTTTAATGCCATG	CTGGCCATTACTTC	1191
Db	1051	TATGCCCTCTATGCCGTCCTGGTCCATGATGGTGGCACTTCTCA	GATGGACATTACTTC	1110
QY	1192	TGCTACATAAAGCTAGCAATGGCCTCTGGTATCAAAATGAAT	GACTCCATTTGATCTTACC	1251
Db	1111	TGTTGTGTCAAAAGCTGTCTATGGGAAGTGGTACAAGATGGAT	GATGATCAACTAAAGTCAC	1170
QY	1252	AGTCATATTAGATCGGTACTCAGCCACACAGCCATATGTCCTTT	TATAT	1301
Db	1171	TGTGATGTGACTTCTGTCCTGAATGAGAANTGCCATATGCTCTT	CTTATGT	1220

RESULT 7

```

US-09-019-095A-21
; Sequence 21, Application US/09019095A
; Patent No. 6287858
; GENERAL INFORMATION:
; APPLICANT: D'Andrea, Alan D.
; APPLICANT: Zhu, Yuan
; TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
; TITLE OF INVENTION: Cell Growth
; FILE REFERENCE: DFCI-435p2A2
; CURRENT APPLICATION NUMBER: US/09/019,095A
; CURRENT FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US96/13884
; PRIOR FILING DATE: 1996-08-07
; PRIOR APPLICATION NUMBER: US 60/002,066
; PRIOR FILING DATE: 1995-08-09
; PRIOR APPLICATION NUMBER: US 60/019,787
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 1523
; TYPE: DNA
; ORGANISM: murine
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(1523)
; OTHER INFORMATION: n = A,T,C or G
US-09-019-095A-21

```

	Query Match	11.8%;	Score 283.6;	DB 3;	Length 1523;
	Best Local Similarity	56.1%;	Pred. No. 3e-79;		
	Matches 585;	Conservative	0;	Mismatches 440;	Indels 18; Gaps 2;
QY	287	CAAAACATACACAAAAAGGATCAAGCCCTTAGTGATGGCATCGTCTCCACAGAAAG	346		
Db	2	CAGCACTATCATCTCTGATGCCCCAGAGCTGCATGTTGAAGCTCAGGTGGTGGAGG	61		
QY	347	TTCTTTTCCCATCTGAGAGATTGTCTTAAGTGGCAACAAATCATAGATGTTGGAGCTG	406		
Db	62	TGCTAACTACCAATGGAAAGTTCAGTCTGAGTTGGGAGAGTCCCTAAGGACACAGGATGCG	121		
QY	407	GCCTCCAGAAATTGGGCAATACCTGTTTTCGCCAATGCAGCATGCGAGTGTTTAACTACA	466		

```

RESULT 8
US-09-280-116-217
; Sequence 217, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176365
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26

```

```

; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 217
; LENGTH: 672
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: ubiquitin carboxyl-terminal hydrolases family 2
; NAME/KEY: misc feature
; LOCATION: (1)..(672)
; OTHER INFORMATION: n = a, t, c or g
US-09-280-116-217

Query Match      8.1%; Score 194.2; DB 4; Length 672;
Best Local Similarity 62.5%; Pred. No. 4.2e-51;
Matches 298; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

QY 301 CAAAGGATCAAGCCCTAGTGATGGATGCTCCACAGAAAGTTCTTTCCCATCT 360
DB 196 CACACGTATGAGAGCTGTGTGACGGAGTCCACAGCCCGCAGAAAGTGTCTTTCCACG 255

QY 361 GAGAAATTTGTTTAAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTG 420
DB 256 GAGCGACTGTCTGAGGTGGGAGCGGGTCTTCGGGTGGCGCAGGACTCCACACCTT 315

QY 421 GGCATATACCTGTTTGGCAATGAGCAGCACTGAGTGTAACTTACATACACACCTTTGCC 480
DB 316 GGCACACCTGCTTCTCTATGCGCACCATCCAGTGTGACCTTACACACACCTCTAGCC 375

QY 481 AATTACATGTCATACATGAACTCCAAACATGTCATGCAAGAGCTTTTGTATGATG 540
DB 376 AACTTACCTGCTCTCCAGGAGCATGCTCGCAGCTGCCACAGGGAAGCTTCTGATCTG 435

QY 541 TGTACATGCAAGACACATATTACCCAGGCACTCAGTAACTCTGGGGAGCTTTATTAACCA 600
DB 436 TGTGTATGAGAACACATGTTCCAGGCTTCGCCAACAGCGGCAAGCCATCAAGCCC 495

QY 601 ATGTTTGTCTCATATGAGATGCGGGGTATAGTAGGACCTCCGTTTGGAAACCAAGAA 660
DB 496 GTCTCTCTTATCCGANAACATGAAAGAGATGCCCGACACTTCCGCTTTGGGAACCAAGAA 555

QY 661 GATGCCATGAAATCTCTCATATACACTGTTGATGCTATGACAGAAAGCATGCTTGAATGGC 720
DB 556 GAGCGCATGAATTCCTCGGGTACACCATGACACACCATGACACCAACCAAGCAATGGC 615

QY 721 AGCAATAAATTAGACAGACACACACCGCCACCACTCTTGTGTCAGATATTTGGA 777
DB 616 TGTGCCAAGTTGGATCTTCAACNNAGCTACTAATTTGGTCCATCAAAATTTTGGGA 672

RESULT 9
US-09-058-489-33
; Sequence 33, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WHI97-08pA
; CURRENT APPLICATION NUMBER: US/09/058,489
; CURRENT FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 10091
; TYPE: DNA
; ORGANISM: Human
; US-09-058-489-34

Query Match      1.9%; Score 44.8; DB 3; Length 10091;
Best Local Similarity 47.5%; Pred. No. 0.0079;
Matches 133; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 857 TGGAGATAAAGGCTGCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTGAGCGGGAAC 916
DB 6895 TGGATATTAGAAATCATCAAAATCTCTTTGACTCTTTTGGAAACAGTATATCAAGAGATT 6954

QY 917 AGCTTGTATGGAGAAACTCGTACAGTGCAGCAAGTGTAAAAAGATGTTCCAGCTTCAA 976
DB 6955 TATTGGAAGGTGCAAAATGCATATCTTTGAAAATGTGATAAAAGGTTGACACAGTAA 7014

QY 977 AGAGGTTTCACTATCATTAGATGCTCTAATGTTTCTTACACTTTCTCTGAAACGTTTTGCA 1036
DB 7015 AGCGCTCTATTAATAAATTTGCTCGGTTCTTGTCTATCCAACTCAACAGGATTTGACT 7074

QY 1037 ATTTTACCGGTGGAAAAATTTGTAAGGATGTGAAATACCTCGAGTATCTTGATATTCGGC 1096
DB 7075 ATGACTGGGAAGAGATGTGCAATTAATTAATGATTTATTTTGAATTTCTTCGAGAGC 7134

QY 1097 CATATATGTCACACCCACGAGAGGCAATTTGTCTAGCT 1136
DB 7135 TGGATATGGGACCTTACACAGTAGCAGGTGTGCAAACT 7174

RESULT 11
US-09-280-116-222
; Sequence 222, Application US/09280116A
```

```

; Patent NO. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 222
; LENGTH: 1998
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: ubiquitin carboxyl-terminal hydrolases family 2
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1998)
; OTHER INFORMATION: n = a, t, c or g
US-09-280-116-222

```

Query Match	1.7%;	Score 42.2;	DB 4;	Length 1998;
Best Local Similarity	50.8%;	Pred. No. 0.017;		
Matches 101;	Conservative 0;	Mismatches 98;	Indels 0;	Gaps 0;
QY	835	GATCCATATCTTCATATAACAATTGGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCATTG	894	
Db	273	GAAGCTTTTCATGGCTCTCAATCTAGGAGTGACTCTTGTGCAGAGTTTGGAAATTTCTTTG	332	
QY	895	GAGCAGTTTTGTGAAGCCGGAACAGCTTGATGGAGAGAAACCTCGTACACAGTCCGACCAAGTGT	954	
Db	333	GACCAATTTTGTAGAGGAGAAGTTCTTAGAAGGAAAGTAATCGCTACTCTGTGAAAAGTGT	392	
QY	955	AAAAAGATGTTCCCGCTTCAAGAGGTTCACTATCCATAGATCCCTTAATGTTCTTTACA	1014	
Db	393	AAAGAAAGAGAAATACAGTGAAGAGGACCTGTATTAAATCTTTACCTAGCGTCTTGGTA	452	
QY	1015	CTTCTCTGAAACGTTTTG	1033	
Db	453	ATTACCTTAATGAGATTGG	471	

RESULT 12
US-09-976-594-862
; Sequence 862, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 862
; LENGTH: 5734
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 1073168.8
US-09-976-594-862

Query Match	1.7%	Score 42.2	DB 4	Length 5734
Best Local Similarity	50.88	Pred. No. 0.035		
Matches 101	Conservative 0	Mismatches 98	Indels 0	Gaps 0

QY	835	GATCCATATCTTGTATATACATTGGAGATAAAGCTGCTCCTCAGAGTCTCAACAGGCAATTG	894
Db	641	GAAGCTTTCATGCTCTCAATCTAGAGTGCATTTCTTGTCAGAGTTGGAAATTTCTTTG	700

QY	895	GAGCAGTTTGTGAAGCCGGAAACAGCTTGATCGAGAAAACTCGTACAAGTCGACGAAGTGT	954
Db			
Db	701	GACCAATTGTTAGAGGAGAAGTCTTAGAAGGAAGTAATCGTACTACTGTGAAAAAGTGT	760
QY	955	AAAAAGATGGTTCGAGTTCCAAAGAGGTTCACTATCCATAGATCCTCTAAATGTTCTTACA	1014
Db			
Db	761	AAAAAAGAGATAACAGTGAAAGACCTGTATTAAATCTTTACCTAGCGTCTTGGA	820
QY	1015	CTTTCICTGAAACGTTTTG	1033
Db			
Db	821	ATTCACCTAATGAGATTG	839

RESULT 13
US-09-976-594-861
; Sequence 861, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 861
; LENGTH: 1683
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 1073168.6
; NAME/KEY: unsure
; LOCATION: 118, 142
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-861

	Query Match	1.7%;	Score 41.8;	DB 4;	Length 1683;
	Best Local Similarity	53.3%;	Pred. No. 0.02;		
	Matches 88;	Conservative 0;	Mismatches 77;	Indels 0;	Gaps 0;
Qy	869	CTGCTCAGAGTGTCAACAAGSCATTTGACGACAGTTTGTGAACCCGGACACGCTTGATGGAG	928		
Db	158	CTTGTCAGAGTTTGGAAATTTCTTTTGACCAATTTGTTAGAGGAGAAAGTTCTACAAGGAA	217		
Qy	929	AAAACCTCGTACAGTCGACGAAGTGTAAGAAAGATGGTTCACAGCTTCAAAGAGGTTCACTA	988		
Db	218	GTAATGGGTACTACTGTGAAAGAGTGTAAGAAAGAGAAATAACAGTCAAAAGACCTGTA	277		
Qy	989	TCCATAGATCCTCTAATGTCTTACACTTTCTCTGAAACGTTTTG	1033		
Db	278	TTAAATCTTACCTACGGCTTGGTAATTCACCTAATGAGATTG	322		

```

RESULT 14
US-08-858-207A-234/c
; Sequence 234, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road

```

us-10-049-745-31.rni

Wed Aug 18 13:53:01 2004

FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6492505 79777CT1
US-09-495-050A-40
Query Match 1 6%; Score 38; DB 4; Length 847;
Best Local Similarity 98.0%; Pred. No. 0.2; Indels 1; Gaps 1;
Matches 49; Conservative 0; Mismatches 0
QY 2363 ATGCGCGCGCAAGCTTATTTCCCTTTAGTGAGGGTTAATTTAGCTTGA 2412
DB 56 ATGCGCGCGCAAGCTTATTTCCCTTTAGTGAGGGTTAATTTAGCTTGA 8
Search completed: August 16, 2004, 20:24:14
Job time : 201 secs

CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA: US/08/858.207A
APPLICATION NUMBER: US/08/858.207A
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 234:
SEQUENCE CHARACTERISTICS:
LENGTH: 730 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-858-207A-234
Query Match 1.6%; Score 38.2; DB 4; Length 730;
Best Local Similarity 49.7%; Pred. No. 0.15;
Matches 97; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
QY 1020 TCTGAACGTTTTCGCAAAATTTACCGGTGGAAAAATGCTAAGATGTGAAATACCTGA 1079
DB 212 TATGAACGTTATGCTGTTTGTAGTGGCGGAGACGACCTGGTATGAACGCTGCTAT 153
QY 1080 GTATCTTGATTCGGCCATATATCTCTCAACCAAGAGAGCAATGCTACGTCTT 1139
DB 152 CCGTGACGTTGTTTCGTCGAAGCAATTCAGAAGGAATGGAAGTGTGTTGATCTATGACGG 93
QY 1140 GTATGCAAGTCTGCTCCACACTGGTTTAAATTCATGCTGCGCAATTAATCTCTGCTACAT 1199
DB 92 ATATGCTGTAAGTGTGCGGTGAAATTCATCCCTAGATGCAAGTTCAGTAGGAGACAT 33
QY 1200 AAAAGCTAGCAATGG 1214
DB 32 CATTTCGTGGTGG 18
RESULT 15
US-09-495-050A-40/c
; Sequence 40, Application US/09495050A
; Patent No. 6492505
; GENERAL INFORMATION:
; APPLICANT: Roopa, Reddy
; APPLICANT: Guegler, Karl, J.
; APPLICANT: Au-Young, Janice
; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED H
; FILE REFERENCE: PA-0013 US
; CURRENT APPLICATION NUMBER: US/09/495,050A
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/118,318
; PRIOR FILING DATE: February 1, 1999
; NUMBER OF SEQ ID NOS: 305
; SOFTWARE: PERL Program
; SEQ ID NO 40
; LENGTH: 847
; TYPE: DNA
; ORGANISM: Homo sapiens

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 16, 2004, 20:24:22 ; Search time 14895 Seconds
(without alignments)
7018.691 Million cell updates/sec

Title: US-10-049-745-31
Perfect score: 2412
Sequence: 1 ggctgtgtgcggcgccgc.....agggttaatttagcttgca 2412

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 30 *ov mpr2*
Total number of hits satisfying chosen parameters: 37

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2412	100.0	2412	6	AX083128 Sequence
2	2321	96.2	4659	6	AX786887 Sequence
3	2287	94.8	2347	6	AX107851 Sequence
4	2265	93.9	3660	6	AX882263 Sequence
5	2265	93.9	3660	6	BD159625 Primer fo
6	2265	93.9	3660	9	BC022759 Homo sapi
7	2259	93.7	3669	9	BC060846 Homo sapi
8	1475	61.2	1679	6	AX364931 Sequence
c 9	722	29.9	152927	9	AC004895 Homo sapi
10	722	29.9	154791	2	AC069393 Homo sapi
c 11	415	17.2	217022	2	AC146146 Pan trogl
12	357	14.8	722	6	AR264045 Sequence
13	253	10.5	217022	2	AC146146
14	245	10.2	5269	6	BD181571 Novel gen
15	232	9.6	521	6	AX867589 Sequence
16	232	9.6	521	6	BD147651 Primer fo
17	205	8.5	457	6	AR264036 Sequence
c 18	205	8.5	154791	2	AC069393 Homo sapi
19	137	5.7	213	6	AX495102 Sequence
c 20	100	4.1	292	6	AX387329 Sequence
21	38	1.6	195906	2	AC136746 Mus muscu
c 22	34	1.4	195906	2	AC136746
c 23	32	1.3	239456	2	AC108995 Rattus no
c 24	32	1.3	270456	2	AC111575 Rattus no
c 25	30	1.2	637	6	AX327423 Sequence
c 26	30	1.2	733	6	AX174745 Sequence
c 27	30	1.2	823	6	AX078266 Sequence
c 28	30	1.2	972	6	AX073683 Sequence
c 29	30	1.2	1106	6	AX135572 Sequence
c 30	30	1.2	1163	6	AX135583 Sequence
31	30	1.2	1487	9	AF121858 Homo sapi
32	30	1.2	1491	6	AX061211 Sequence
33	30	1.2	1702	6	AX073668 Sequence
34	30	1.2	1950	6	AX081430 Sequence
c 35	30	1.2	1954	6	AX077793 Sequence
c 36	30	1.2	1987	6	BD260129 Matrix-re
c 37	30	1.2	3465	6	BD224624 Protein t

ALIGNMENTS

RESULT 1	AX083128	AX083128	2412 bp	DNA	linear	PAT 28-FEB-2001
LOCUS	Sequence 31 from Patent WO0110903.					
DEFINITION	Sequence 31 from Patent WO0110903.					
ACCESSION	AX083128					
VERSION	AX083128.1	GI:13185035				
KEYWORDS	Homo sapiens (human)					
SOURCE	Homo sapiens					
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
REFERENCE	1					
AUTHORS	Yue, H., Lal, P., Tang, Y.T., Bandnan, O., Baughn, M.R., Azimzai, Y., Lu, D.A. and Yang, J.					
TITLE	Proteases and protease inhibitors					
JOURNAL	Patent: WO 0110903-A 31 15-FEB-2001;					
FEATURES	Incyte Genomics, Inc. (US)					
source	Location/Qualifiers					
	1. .2412					
	/organism="Homo sapiens"					
	/mol_type="unassigned DNA"					
	/db_xref="taxon:9606"					
	/note="Incyte ID No: 1299481CB1"					

ORIGIN

Query Match					
Best Local Similarity		100.0%; Score 2412; DB 6; Length 2412;			
Matches 2412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	GGGCTGTGTGGCGGCGGCGGCGGCGGCGAGGGGGATGAGCGAGCGCGAGCCGGGT	60		
Dd	1	GGGCTGTGTGGCGGCGGCGGCGGCGGCGAGGGGGATGAGCGAGCGCGAGCCGGGT	60		
QY	61	CAGAGTTGAACAATGACCATAGTGTGACAAAGCTTCTGAATCTTCAGACCCTCAGCCTAT	120		
Dd	61	CAGAGTTGAACAATGACCA TAGITGACAAAGCTTCTGAATCTTCAGACCCTCAGCCTAT	120		
QY	121	CAGAATCAGCCTGGCAGCTCCGAGGCAGTGCTCACCTGGAGACATGGATGAGTCTTGCC	180		
Dd	121	CAGAATCAGCCTGGCAGCTCCGAGGCAGTCTCACCTGGAGACATGGATGAGTCTTGCC	180		
QY	181	AGCTGGGGTGCTGTGCTTTCATTGAATGATGTCGTAATCACAACACTTCTTTAGAACCA	240		
Dd	181	AGCTGGGGTGCTGTGCTTTCATTGAATGATGTCGTAATCACAACACTTCTTTAGAACCA	240		
QY	241	GTACCTGGTGCTGTAGTTTATTCGAGTTCATCTGTGTACTGTATAAATCAAACCATCACCA	300		
Dd	241	GTACCTGGTGCTGTAGTTTATTCGAGTTCATCTGTACTGTATAAATCAAACCATCACCA	300		
QY	301	CAAAAGGATCAAGCCCTAGTGTGATGTCATCGCTCTCCAAGAAAGTCTTTTCCCATCT	360		
Dd	301	CAAAAGGATCAAGCCCTAGTGTGATGTCATCGCTCTCCAAGAAAGTCTTTTCCCATCT	360		
QY	361	GAGAGATTGTCTTAAATGGCAACAACATCATAGATTGGAGCTGGGTCCAGAAATTTG	420		
Dd	361	GAGAAGATTGTCTTAAATGGCAACAACATCATAGATTGGAGCTGGGTCCAGAAATTTG	420		
QY	421	GGCAATACCTGTTTTGCCAATGCGAGCTGCAAGTGTTHAACTACACCACTCTTGCC	480		
Dd	421	GGCAATACCTGTTTTGCCAATGCGAGCTGCAAGTGTTHAACTACACCACTCTTGCC	480		
QY	481	AATTCATGCTATCATCATGAACACTCCAAAACATGTCATGCAAGAGGCTTTGTATGATG	540		
Dd	481	AATTCATGCTATCATCATGAACACTCCAAAACATGTCATGCAAGAGGCTTTGTATGATG	540		
QY	541	TGTAAATGCAAGACATATTAACCAAGCACTCAGTAATCTCTGGGGAGCTTATTAACA	600		
Dd	541	TGTAAATGCAAGACATATTAACCAAGCACTCAGTAATCTCTGGGGAGCTTATTAACA	600		
QY	601	ATGTTTGTCAATCAATGAGATCGGCGTATAGTAGGCACTCCGTTTGGAAAACAGAA	660		
Dd	601	ATGTTTGTCAATCAATGAGATCGGCGTATAGTAGGCACTCCGTTTGGAAAACAGAA	660		
QY	661	GATGCCANTGAATTCCTTCAATACACTGTTGATGCTATGCAAGAGCAATGCTGAATGCC	720		
Dd	661	GATGCCANTGAATTCCTTCAATACACTGTTGATGCTATGCAAGAGCAATGCTGAATGCC	720		
QY	721	AGCAATAAATPAGACAGACACACCCAGGCCACCACTCTTGTGTCAGATTTGGAGGA	780		
Dd	721	AGCAATAAATPAGACAGACACACCCAGGCCACCACTCTTGTGTCAGATTTGGAGGA	780		
QY	781	TACCTAAGACTAGAGTCAATGTTTAAATGCAAGGCGGTTTCAGATACCTTTTGAATCCA	840		
Dd	781	TACCTAAGACTAGAGTCAATGTTTAAATGCAAGGCGGTTTCAGATACCTTTTGAATCCA	840		
QY	841	TATCTCATATACATTTGGAGATAAAGGCTGCTCAGAGTGTCAACAAAGGCAATGGAGCAG	900		
Dd	841	TATCTCATATACATTTGGAGATAAAGGCTGCTCAGAGTGTCAACAAAGGCAATGGAGCAG	900		
QY	901	TTTGTGAAGCGGGAACAGCTTGTATGGAGAAAACTCGTACAAGTGCACAAGTGTAAAG	960		
b	901	TTTGTGAAGCGGGAACAGCTTGTATGGAGAAAA CTGTAAGTGCACAAGTGTAAAG	960		
QY	961	ATGTTTCCAGCTTCAAGAGGTTCACTATCCATAGATCTCTTAATGTTCTTACACTTCT	1020		
b	961	ATGTTTCCAGCTTCAAGAGGTTCACTATCCATAGATCTCTTAATGTTCTTACACTTCT	1020		

QY	1021	CTGAAACGTTTGTGCAAAATTTTACGGTGGAAAAATTTGCTAAAGATGTGAAATACCCCTGAG	1080
DB	1021	CTGAAACGTTTGTGCAAAATTTTACGGTGGAAAAATTTGCTAAAGATGTGAAATACCCCTGAG	1080
QY	1081	TATCTTGATATTCGGCCATATATGTCCTCAACCCGAGAGCCCAATTGTCTACGCTCTTG	1140
DB	1081	TATCTTGATATTCGGCCATATATGTCCTCAACCCGAGAGCCCAATTGTCTACGCTCTTG	1140
QY	1141	TATCAGTGTGCTGCCACACTGGTTTTAATTTGCCATGCTGGCCATCTACTCTGCTACACATA	1200
DB	1141	TATCAGTGTGCTGCCACACTGGTTTTAATTTGCCATGCTGGCCATCTACTCTGCTACACATA	1200
QY	1201	AAAGCTAGCAATGGCTCTGCTGATCAAATGAATGACTCCATTTGATCTACCAAGTATAT	1260
DB	1201	AAAGCTAGCAATGGCTCTGCTGATCAAATGAATGACTCCATTTGATCTACCAAGTATAT	1260
QY	1261	AGATCGGTACTCA3CCCAACAGCCTATGTGCTCTTTATATACAGTCCCATGATGTGAAA	1320
DB	1261	AGATCGGTACTCA3CCCAACAGCCTATGTGCTCTTTATATACAGTCCCATGATGTGAAA	1320
QY	1321	AAATGGAGGTGAACCTTACTCATCCACCATAGCCCCCGGCCAGTCTCTCCCGCCCGCTC	1380
DB	1321	AAATGGAGGTGAACCTTACTCATCCACCATAGCCCCCGGCCAGTCTCTCCCGCCCGCTC	1380
QY	1381	ATCAGTCCAGCGGTGTGTCACCAACAAACAGCGCTGCGCCAGGCTTTATCGGACCAAGTT	1440
DB	1381	ATCAGTCCAGCGGTGTGTCACCAACAAACAGCGCTGCGCCAGGCTTTATCGGACCAAGTT	1440
QY	1441	CCCTCTCAATGATAAAGAAATCCACCTCACTTAAATGGGACTGGACCAATCGAAAGACAG	1500
DB	1441	CCCTCTCAATGATAAAGAAATCCACCTCACTTAAATGGGACTGGACCAATCGAAAGACAG	1500
QY	1501	CCAAAGCAGTTCCAATGTCGAGTCTTAACGGAAATTCAGTGTCAACAGGGCTAGTCTCTGT	1560
DB	1501	CCAAAGCAGTTCCAATGTCGAGTCTTAACGGAAATTCAGTGTCAACAGGGCTAGTCTCTGT	1560
QY	1561	AAATGCTTCAGCTTCTGTCCAAAACCTGGTCAAGTTAAATAGTGTCTTCAAGTATCCCAAGAACAT	1620
DB	1561	AAATGCTTCAGCTTCTGTCCAAAACCTGGTCAAGTTAAATAGTGTCTTCAAGTATCCCAAGAACAT	1620
QY	1621	CCTAAGAACAAATAATTAATCAATCAGTATTCACAAAGTTGCTGTTCGCCAGTGTCTAG	1680
DB	1621	CCTAAGAACAAATAATTAATCAATCAGTATTCACAAAGTTGCTGTTCGCCAGTGTCTAG	1680
QY	1681	TCTCAACCTTAACCTTCATAGTAATTTCTTTGGAGAACCTTACCAAGCCCGTTCCTCTCTCT	1740
DB	1681	TCTCAACCTTAACCTTCATAGTAATTTCTTTGGAGAACCTTACCAAGCCCGTTCCTCTCTCT	1740
QY	1741	ACCATTACCAATTCCTGAGTACAGTCTACCTCGAACGCATCTACGATGTCAAGTTCTAGT	1800
DB	1741	ACCATTACCAATTCCTGAGTACAGTCTACCTCGAACGCATCTACGATGTCAAGTTCTAGT	1800
QY	1801	AAAAGTAAACAAAACCGATCCCGCGCAGTGAATCCTGCTCCCGCCCGTGTGATGCGCAA	1860
DB	1801	AAAAGTAAACAAAACCGATCCCGCGCAGTGAATCCTGCTCCCGCCCGTGTGATGCGCAA	1860
QY	1861	TCCAAAGCTGAATCTCAGCGTGTGTCCTATGCGCGCAGTCTCTCAGGACTCTGAC	1920
DB	1861	TCCAAAGCTGAATCTCAGCGTGTGTCCTATGCGCGCAGTCTCTCAGGACTCTGAC	1920
QY	1921	GAGGAGTCAAAAGGG3CTGGGCAAGAGAAATGGGATTTGGTACGATTTGAGCTCCCACTCT	1980
DB	1921	GAGGAGTCAAAAGGG3CTGGGCAAGAGAAATGGGATTTGGTACGATTTGAGCTCCCACTCT	1980
QY	1981	CCCGGCCAAGATGCTGAAGATGAGGAGGCCACTCCGACAGGCTTCAAGAACCCCATGACC	2040
DB	1981	CCCGGCCAAGATGCTGAAGATGAGGAGGCCACTCCGACAGGCTTCAAGAACCCCATGACC	2040
QY	2041	CTAAACCGTGTCTAATAGTTCAGACAGCGACAGTACCAGAAAGAAACGGCTTAGCGCT	2100
DB	2041	CTAAACCGTGTCTAATAGTTCAGACAGCGACAGTACCAGAAAGAAACGGCTTAGCGCT	2100

```
QY 2101 GATGGTCCAGCTGCCAAGCGCAGCTGCCCTGCACCTCAGAAAATCCCTTTGTAAGGCA 2160
Db |||||
QY 2101 GATGGTCCAGCTGCCAAGCGCAGCTGCCCTGCACCTCAGAAAATCCCTTTGTAAGGCA 2160
Db |||||
QY 2161 AACGGTCTTCCTGAAAGTTGATGCTGCTCCTTTGCTGTCTCTCCAGAAAGCAAAATC 2220
Db |||||
QY 2161 AACGGTCTTCCTGAAAGTTGATGCTGCTCCTTTGCTGTCTCTCCAGAAAGCAAAATC 2220
Db |||||
QY 2221 TTAGAGACCTTCAGGCTTAGCAACAACTGAAAGCTCGACGGATGAATGAGTGCACCT 2280
Db |||||
QY 2221 TTAGAGACCTTCAGGCTTAGCAACAACTGAAAGCTCGACGGATGAATGAGTGCACCT 2280
Db |||||
QY 2281 GGAGCAGAGAGGGGCCCTCCCGAGGACCGGACCGCCAGCTCAGCTGGCAGCCCGGCC 2340
Db |||||
QY 2281 GGAGCAGAGAGGGGCCCTCCCGAGGACCGGACCGCCAGCTCAGCTGGCAGCCCGGCC 2340
Db |||||
QY 2341 GCCGAATCCCTGAGGAGCCAGATGCGGCGCGCAAGCTTATCCCTTTAGTGGGGTTAA 2400
Db |||||
QY 2341 GCCGAATCCCTGAGGAGCCAGATGCGGCGCGCAAGCTTATCCCTTTAGTGGGGTTAA 2400
Db |||||
QY 2401 TTTTAGCTTGCA 2412
Db |||||
QY 2401 TTTTAGCTTGCA 2412
Db |||||

RESULT 2
LOCUS AX786887 4659 bp DNA linear PAT 17-JUL-2003
DEFINITION Sequence 22 from Patent WO0238744.
ACCESSION AX786887
KEYWORDS AX786887.1 GI:32954188
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Lee,E.A., Hafalia,A.J., Yue,H., Lal,P.G., Yao,M.G., Lu,Y.,
Chawla,N.K., Warren,B.A., Lu,D.A., Baughn,M.R., Delegeane,A.M.,
Burford,N., Borowsky,M.L., Lee,S., Xu,F., Griffin,J.A.,
Kallick,D.A., Gandhi,A.R., Arvizu,C.S., Ison,C.H., Tang,Y.T.,
Azimzai,Y., Elliott,V.S., Swarnakar,A., Ramkumar,J., Nguyen,D.B.,
Tribouley,C.M., Lo,T.P., Au-Young,J., Thangavelu,K. and Kearney,L.
Proteases
TITLE Patent: WO 0238744-A 22 16-MAY-2002;
JOURNAL Incyte Genomics, Inc. (US)
FEATURES
source
location/Qualifiers
1..4659
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/noe="Incyte ID No: 7997881CB1"

Query Match 96.2%; Score 2321; DB 6; Length 4659;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2371; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGCTGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 60
Db |||||
QY 83 GGGCTGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 142
Db |||||
QY 61 CAGAGTTGAACCAATGACCAATGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTAT 120
Db |||||
QY 143 CAGAGTTGNACCAATGACCAATGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTAT 202
Db |||||
QY 121 CAGAAATCAGCTGGCAGCTCCGAGGCGAGTCTCACCTGGAGACATGATGCGAGTTCTGCC 180
Db |||||
QY 203 CAGAAATCAGCTGGCAGCTCCGAGGCGAGTCTCACCTGGAGACATGATGCGAGTTCTGCC 262
Db |||||
QY 181 AGCTGGGTGCTGTCTTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Db |||||
QY 263 AGCTGGGTGCTGTCTTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 322
Db |||||
```

```
QY 241 GTAACCTGGTGTCTAGTATTTATTCGAGTTCACTCTGTACCTGTATAAATCAAAACCAATCACA 300
Db |||||
QY 323 GTAACCTGGTGTCTAGTATTTATTCGAGTTCACTCTGTACCTGTATAAATCAAAACCAATCACA 382
Db |||||
QY 301 CAAAGAGATCAAGCCCTAGGTGATGGCATGGCTCTCTCCACAGAAAGTTCTTTTCCCAATCT 360
Db |||||
QY 383 CAAAGAGATCAAGCCCTAGGTGATGGCATGGCTCTCTCCACAGAAAGTTCTTTTCCCAATCT 442
Db |||||
QY 361 GAGAAGATTTGTCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTG 420
Db |||||
QY 443 GAGAAGATTTGTCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTG 502
Db |||||
QY 421 GGAATACCTGTTTGTGCAATGCAAGCACTGCAAGTGTAACTTACCTACACACCACTCTTTGCC 480
Db |||||
QY 503 GGAATACCTGTTTGTGCAATGCAAGCACTGCAAGTGTAACTTACCTACACACCACTCTTTGCC 562
Db |||||
QY 481 AATTACATGCTATCAGATGAACACTCCAAACATGTCTATGCAAGAGCTTTTGTATGATG 540
Db |||||
QY 563 AATTACATGCTATCAGATGAACACTCCAAACATGTCTATGCAAGAGCTTTTGTATGATG 622
Db |||||
QY 541 TGTACAAATGCAAGCACAATATTACCAGGCACTCAGTAATCTCTGGGACGTTTATTAACCA 600
Db |||||
QY 623 TGTACAAATGCAAGCACAATATTACCAGGCACTCAGTAATCTCTGGGACGTTTATTAACCA 682
Db |||||
QY 601 ATGTTTGTCTATCAATGAGATGCGGCGGTATAGTAGGCACTCTCCGTTTGGAAACCAAGAA 660
Db |||||
QY 683 ATGTTTGTCTATCAATGAGATGCGGCGGTATAGTAGGCACTCTCCGTTTGGAAACCAAGAA 742
Db |||||
QY 661 GATGCCCAGTAATTCCTTCAATACACTGTTGATGCTATGCAAGAGCTTTTGTATGATG 720
Db |||||
QY 743 GATGCCCAGTAATTCCTTCAATACACTGTTGATGCTATGCAAGAGCTTTTGTATGATG 802
Db |||||
QY 721 AGCAATAAATTAGACAGACACACCCAGGCGCCACCTCTCTGTTGTGAGATTTTGGAGGA 780
Db |||||
QY 803 AGCAATAAATTAGACAGACACACCCAGGCGCCACCTCTCTGTTGTGAGATTTTGGAGGA 862
Db |||||
QY 781 TACCTAAGATCTAGAGTCAAAATGTTTAAATGCAAGGCGGTTTTCAGATCTTTTGTATCCA 840
Db |||||
QY 863 TACCTAAGATCTAGAGTCAAAATGTTTAAATGCAAGGCGGTTTTCAGATCTTTTGTATCCA 922
Db |||||
QY 841 TATCTTGTATTAACATTTGAGATAAAGGCTGCTCAGAGTGTCAACAGGCAATTTGGAGCAG 900
Db |||||
QY 923 TATCTTGTATTAACATTTGAGATAAAGGCTGCTCAGAGTGTCAACAGGCAATTTGGAGCAG 982
Db |||||
QY 901 TTTGTGAGCGCGGAAACAGCTTGTATGAGAGAAAATCTCTACAGTGCAGCAAGTCTTAAAG 960
Db |||||
QY 983 TTTGTGAGCGCGGAAACAGCTTGTATGAGAGAAAATCTCTACAGTGCAGCAAGTCTTAAAG 1042
Db |||||
QY 961 ATGGTTCCAGCTTCAAGAGGTTCACTATCCATAGATCTCTTAATGTTCTTACACTTTCT 1020
Db |||||
QY 1043 ATGGTTCCAGCTTCAAGAGGTTCACTATCCATAGATCTCTTAATGTTCTTACACTTTCT 1102
Db |||||
QY 1021 CTGAAACGTTTGTGCAAAATTTTACCGGTGGAATAATGCTTAAGATGTGAAATACCTGAG 1080
Db |||||
QY 1103 CTGAAACGTTTGTGCAAAATTTTACCGGTGGAATAATGCTTAAGATGTGAAATACCTGAG 1162
Db |||||
QY 1081 TATCTTGTATTTGCGGCATATATGTCTCAACCAACGAGAGCCAAATCTCTACAGTCTTTG 1140
Db |||||
QY 1163 TATCTTGTATTTGCGGCATATATGTCTCAACCAACGAGAGCCAAATGCTTACAGTCTTTG 1222
Db |||||
QY 1141 TATGAGTGTGCTGCTCCACACTGCTTTTAAATGCCATGTGGCCATTACTTCTGTACATA 1200
Db |||||
QY 1223 TATGAGTGTGCTGCTCCACACTGCTTTTAAATGCCATGTGGCCATTACTTCTGTACATA 1282
Db |||||
QY 1201 AAAGTACCAATGGCCTCTGCTATCAAAATGATGATGATGATGATGATGATGATGATGAT 1260
Db |||||
QY 1283 AAAGTACCAATGGCCTCTGCTATCAAAATGATGATGATGATGATGATGATGATGATGAT 1342
Db |||||
QY 1261 AGATCGGTACTCAGCCAAACAGCCCTATGTGCTCTTTTATATCAGGTCCCATGATGAAA 1320
Db |||||
QY 1343 AGATCGGTACTCAGCCAAACAGCCCTATGTGCTCTTTTATATCAGGTCCCATGATGAAA 1402
Db |||||
```



```
QY 1321 AATGGAGGTGAACCTTACTCATCCACCATTAGCCCGGCGAGTCCCTCTCCCGCCCGGTC 1380
|
|
|
Db 1403 AATGGAGGTGAACCTTACTCATCCACCATTAGCCCGGCGAGTCCCTCTCCCGCCCGGTC 1462
|
|
|
QY 1381 ATCAGTCAGCGGTTGTGTCACCAACAAACAGGCTGCGCAGGCTTTATCGGACACACAGCTT 1440
|
|
|
Db 1463 ATCAGTCAGCGGTTGTGTCACCAACAAACAGGCTGCGCAGGCTTTATCGGACACACAGCTT 1522
|
|
|
QY 1441 CCTCTCACATGATAAAGATCCACCTTCACTTAAATGGAGTGGAGTGGACCAATGAAAGACAG 1500
|
|
|
Db 1523 CCTCTCACATGATAAAGATCCACCTTCACTTAAATGGAGTGGAGTGGACCAATGAAAGACAG 1582
|
|
|
QY 1501 CCAAGCAGTTCATGTCCAGTCCTAACGGGAATTCAGTTCACAGGCTAGTCCGTT 1560
|
|
|
Db 1583 CCAAGCAGTTCATGTCCAGTCCTAACGGGAATTCAGTTCACAGGCTAGTCCGTT 1642
|
|
|
QY 1561 AATGCTTCAGTCTGTCCAAACCTGTCTAGTTAATAGTTCCTCAGTGTATCCCAAGACAT 1620
|
|
|
Db 1643 AATGCTTCAGTCTGTCCAAACCTGTCTAGTTAATAGTTCCTCAGTGTATCCCAAGACAT 1702
|
|
|
QY 1621 CCTAAGAAACAAAAATTAACATCAGTATTCACAAAGTTGCTGTTCGCGAGTGTGAC 1680
|
|
|
Db 1703 CCTAAGAAACAAAAATTAACATCAGTATTCACAAAGTTGCTGTTCGCGAGTGTGAC 1762
|
|
|
QY 1681 TCTCAACCTTAACCTTCATAGTAATCTTTGGAGAACCTCAAGCCGCTTCCCTCTTCT 1740
|
|
|
Db 1763 TCTCAACCTTAACCTTCATAGTAATCTTTGGAGAACCTCAAGCCGCTTCCCTCTTCT 1822
|
|
|
QY 1741 ACCATTACCAATTTGCGAGTACAGTCTACCTCGAAGCATCTACGATGTCTAGTTTCTAGT 1800
|
|
|
Db 1823 ACCATTACCAATTTGCGAGTACAGTCTACCTCGAAGCATCTACGATGTCTAGTTTCTAGT 1882
|
|
|
QY 1801 AAAGTAAACAAACCGATCCCGGAGTGAATCTGCTCCCAAGCCGCTGATGAATGGCAAA 1860
|
|
|
Db 1883 AAAGTAAACAAACCGATCCCGGAGTGAATCTGCTCCCAAGCCGCTGATGAATGGCAAA 1942
|
|
|
QY 1861 TCGAAGCTGAATCTCAGGCTGTGTCCTATGCGCGGAGTCTCTGAGACTCTGAC 1920
|
|
|
Db 1943 TCGAAGCTGAATCTCAGGCTGTGTCCTATGCGCGGAGTCTCTGAGACTCTGAC 2002
|
|
|
QY 1921 GAGGAGTCAAAAGGGCTGGGCAAGGAGATGGGATTTGGTACGATGTGAGCTTCCCACTCT 1980
|
|
|
Db 2003 GAGGAGTCAAAAGGGCTGGGCAAGGAGATGGGATTTGGTACGATGTGAGCTTCCCACTCT 2062
|
|
|
QY 1981 CCGGCCCAAGATGCGCAAGATGAGGAGCCATCTCGGACGAGCTTCAAGAACCCATGACC 2040
|
|
|
Db 2063 CCGGCCCAAGATGCGCAAGATGAGGAGCCATCTCGGACGAGCTTCAAGAACCCATGACC 2122
|
|
|
QY 2041 CTAAACGGTGTCTAATAGTCAGACAGCAGCAGTACCGAAGAAACGGCCTAGCGCT 2100
|
|
|
Db 2123 CTAAACGGTGTCTAATAGTCAGACAGCAGCAGTACCGAAGAAACGGCCTAGCGCT 2182
|
|
|
QY 2101 GATGTGCGCAGTGCACAGCGCAGCTGCGCTGCACTCAGAAATCCCTTTGCTAAGGCA 2160
|
|
|
Db 2183 GATGTGCGCAGTGCACAGCGCAGCTGCGCTGCACTCAGAAATCCCTTTGCTAAGGCA 2242
|
|
|
QY 2161 AACGCTTTCCTGGAAAGTGTATGCTGCTCTTCTGCTGCTCTCCGAGAACCAAAATC 2220
|
|
|
Db 2243 AACGCTTTCCTGGAAAGTGTATGCTGCTGCTCTTCTGCTGCTCTCCGAGAACCAAAATC 2302
|
|
|
QY 2221 TTAGAGACCTTCAGGCTTAGCAACAACTGAAAGGCTTCAGCGATGAAATGAGTGCACCT 2280
|
|
|
Db 2303 TTAGAGACCTTCAGGCTTAGCAACAACTGAAAGGCTTCAGCGATGAAATGAGTGCACCT 2362
|
|
|
QY 2281 GGAGCAGAGAGGGGCCCTCCGAGGACCGGACGCGAGCCTCAGCCTGGCAGCCCCGCC 2340
|
|
|
Db 2363 GGAGCAGAGAGGGGCCCTCCGAGGACCGGACGCGAGCCTCAGCCTGGCAGCCCCGCC 2422
|
|
|
QY 2341 GCCGAATCCCTGGAGGAGCAGATGCGGCGC 2372
|
|
|
Db 2423 GCCGAATCCCTGGAGGAGCAGATGCGGCGC 2454
|
|
|
```

```
RESULT 3
AX107851
LOCUS 2347 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 2 from Patent WO0123589.
ACCESSION AX107851
VERSION AX107851.1 GI:13923242
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Kapeller-Lieberman,R.
TITLE 23431, a novel human ubiquitin protease
JOURNAL Patent: WO 0123589-A 2 05-APR-2001; Millennium Pharmaceuticals, Inc. (US)
FEATURES
source
1. 2347
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/notice="23431 Ubiquitin protease"
61..>2347
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC37760.1"
/db_xref="GI:13923243"
/db_xref="REMBEL:CAC37760"
/translation="MTIVKASSESDPSAYONOPSSSEAVSPGDMAGSAGSAGVSSLL
NDVSNHTLSLGPVPGAVVSYSSVPDKSPKSPQKDALGDIAPKPVLPSEKICLK
WOOTHRVAGLQNLNCTCFANAALOCITPTPLANYMLSHEHSTCHAGFCMCTMQ
AHIQTALNSPDGVIKPMFVINEMRIARHFRGNOEDAHFLETITLIEKAAQSVKALEQ
KLDRHTQATTLVCOIFGGYLRSRVKLNCKGSDTFDPLDITLIEKAAQSVKALEQ
FVKEQDNGNSYKCKKMPKSKPTIHRSSNVLTLSLRPANFTGGKIAKDVKY
PBYLDIRPMSQPNGBPIVYLVAVLHVTGNCAGHVCVFKASNGIWMYQNDISVS
TSDIRLSQOAYVLFYIRSHDVKNKGELTHPTSPGQSSPRVTSORVWTKQAAPG
FIGQLPSHMIKPNHPLNGTLPDTPSSSSPNNGNSVNRASPVNRASVQWNSVN
RSSVPIHPKQKITISIHNLKPVRCQSPQNLHNSLENFTKVPFSTITNSAVQST
SNAGTMSVSSKVTPIPRSSCSQPVNMGSKLNSVLVYPYGAESSESDLESKGLGK
ENGIGTIVSSHPGQDAEDBATPHELOEPMTLNGANSADSDPKENGLAPDGASQC
GQPAIHSENPFAKANGLPGLKMPAPLLSLPEDKILETLFRLSNKLKGSTDEMSAPGAER
GPPEDRDHAEPQPGSPAAESLEE"
ORIGIN
Query Match 94.8%; Score 2287; DB 6; Length 2347;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2337; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 22 GCGGCGCGGAGGGGAGTGGAGCGAGCGCGGCGGCTCAGAGTTGAACATGACCATA 81
Db 10 GCGGCGCGGAGGGGAGTGGAGCGAGCGCGGCTCAGAGTTGAACATGACCATA 69
QY 82 GTTGCAAAAGCTTCGAATCTTCAGACCCATCAGCTATCAGATCGAGCTCC 141
Db 70 GTTGCAAAAGCTTCGAATCTTCAGACCCATCAGCTATCAGATCGAGCTCC 129
QY 142 GAGGCGAGTCTCACCTGGAGACATGGATGCGAGCTTCTGCCAGCTGGGGTCTGTCTTCA 201
Db 130 GAGGCGAGTCTCACCTGGAGACATGGATGCGAGTCTGCCAGCTGGGGTCTGTCTTCA 189
QY 202 TTGAATGATGTCTCAAAATCACACATCTTCTTTAGACCAAGTACCTGGTCTGTCTTAT 261
Db 190 TTGAATGATGTCTCAAAATCACACATCTTCTTTAGACCAAGTACCTGGTCTGTCTTAT 249
QY 262 TCGAGTTTCATCTGTACCTGATATAATCAAAACCATCAACAAAGGATCAAGCCCTAGT 321
Db 250 TCGAGTTTCATCTGTACCTGATATAATCAAAACCATCAACAAAGGATCAAGCCCTAGT 309
QY 322 GATGCGATCGCTCCCTCCACAGAAAGTCTTTTCCCATCTGAGAAATTTGCTTTAAGTGG 381
Db 310 GATGCGATCGCTCCCTCCACAGAAAGTCTTTTCCCATCTGAGAAATTTGCTTTAAGTGG 369
QY 382 CAACAAACATCAGTAGTGTGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTTGGCAAT 441
```

370 CAAACAACCTCATAGAGTTGGAGTGGGCTCCAGAAATTTGGCCATACCTGTTTGGCAAT 429
442 GCAGCAGCTGAGTGTAACTTACCTTACACACACCTCTGGGCCAAATTAATGCTATCACATGAA 501
430 GCAGCAGCTGAGTGTAACTTACCTTACACACACCTCTGGGCCAAATTAATGCTATCACATGAA 489
502 CACTCCAAAACATGTCATGCAGAAAGGCTTTTGTATGATGTGTACAAATGCAAGCAGCATATT 561
490 CACTCCAAAACATGTCATGCAGAAAGGCTTTTGTATGATGTGTACAAATGCAAGCAGCATATT 549
562 ACCAGGCACTCAGTAATCTCGGGGACGTTTATTAACCAATGTTTGTCAATCAATGAGATG 621
550 ACCAGGCACTCAGTAATCTCGGGGACGTTTATTAACCAATGTTTGTCAATCAATGAGATG 609
622 CGCGGTATAGCTAGGCACTCGCTTTTGGAAACCAAGAGATGCCCATGAATTCCTTCAA 681
610 CGCGGTATAGCTAGGCACTTCGCTTTTGGAAACCAAGAGATGCCCATGAATTCCTTCAA 669
682 TACACTGTTGATGCTATGCAGAAAGCATCTTGAATGGCAGCAATTAATTAAGCAGACAC 741
670 TACACTGTTGATGCTATGCAGAAAGCATCTTGAATGGCAGCAATTAATTAAGCAGACAC 729
742 ACCAGGCCACCACTCTTGTGTCAGATATTTGGAGGATACCTAAGATCTTAGAGTCAAA 801
730 ACCAGGCCACCACTCTTGTGTCAGATATTTGGAGGATACCTAAGATCTTAGAGTCAAA 789
802 TGTTTAAATGCAAGGCGCTTTCAGATACCTTTTGTATCCATATCTTGAATTAACATTTGGAG 861
790 TGTTTAAATGCAAGGCGCTTTCAGATACCTTTTGTATCCATATCTTGAATTAACATTTGGAG 849
862 ATAAAGGCTGCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTGAAGCCGGAACAGCTT 921
850 ATAAAGGCTGCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTGAAGCCGGAACAGCTT 909
922 GATGAGAGAAACTCGTACAAAGTCGACAAAGTGTAAAGATGGTCCAGCTTCAAGAGG 981
910 GATGAGAGAAACTCGTACAAAGTCGACAAAGTGTAAAGATGGTCCAGCTTCAAGAGG 969
982 TTCATATCCATAGATCTCTTAATGTTCTTACATCTTCTCTGAAGCTTTTGAATTTT 1041
970 TTCATATCCATAGATCTCTTAATGTTCTTACATCTTCTCTGAAGCTTTTGAATTTT 1029
1042 ACCGTGTGAAAAATGCTAAGAGTGTAAATACCTCGATGATCTTGTATATCGGCATAT 1101
1030 ACCGTGTGAAAAATGCTAAGAGTGTAAATACCTCGATGATCTTGTATATCGGCATAT 1089
1102 ATGTCTCAACCCACGAGAGCAATTTGTCTAGCTCTTGTATGCAAGTGTGGTCCACACT 1161
1090 ATGTCTCAACCCACGAGAGCAATTTGTCTAGCTCTTGTATGCAAGTGTGGTCCACACT 1149
1162 GGTTTTAATGTCATGCTGGCCATTAATCTTCTGTACATAAAGCTAGCAATGGCCCTCTGG 1221
1150 GGTTTTAATGTCATGCTGGCCATTAATCTTCTGTACATAAAGCTAGCAATGGCCCTCTGG 1209
1222 TATCAAAATGAATGACTCCATCTGATCTACAGTATATTAGATCGTACTCAGCCAAACAA 1281
1210 TATCAAAATGAATGACTCCATCTGATCTACAGTATATTAGATCGTACTCAGCCAAACAA 1269
1282 GCCTATGTGCTCTTTTATATCAGGTCCTCATGATGTGAAAAATGGAGGTGAACTTACTCAT 1341
1270 GCCTATGTGCTCTTTTATATCAGGTCCTCATGATGTGAAAAATGGAGGTGAACTTACTCAT 1329
1342 CCCACCATAGCCCGCCGAGCTCTCTCCCGCCGCTCATCAGTCAGCGGTTGTCCAC 1401
1330 CCCACCATAGCCCGCCGAGCTCTCTCCCGCCGCTCATCAGTCAGCGGTTGTCCAC 1389
1402 AACAAACAGGCTGCGCAGGCTTTATCGACACACAGCTTCCCTCTCACATGATAAGAT 1461
1390 AACAAACAGGCTGCGCAGGCTTTATCGACACACAGCTTCCCTCTCACATGATAAGAT 1449
1462 CCACCTCCTTAATGGGACTGGACCATTTGAAAGACACGCCCAAGAGTTCCCATGTGCGCT 1521

1450 CCACCTCACTTAATCGGACTGGACCATTTGAAAGACACGCCCAAGCAGTTCCTATGTGAGT 1509
1522 CTTAACGGGAATTTCCAGTGTCAACAGGGCTAGTCTCTTAAATGCTTTCAGCTTCTGTGCCAA 1581
1510 CTTAACGGGAATTTCCAGTGTCAACAGGGCTAGTCTCTTAAATGCTTTCAGCTTCTGTGCCAA 1569
1582 AACTGCTCAGTTAATAGTCTCCTCAGTGATCCCAAGACATCTTAAGAAACAAAAATTACA 1641
1570 AACTGCTCAGTTAATAGTCTCCTCAGTGATCCCAAGACATCTTAAGAAACAAAAATTACA 1629
1642 ATCAGTATTCAACAACAGTTTGCCTGTTGCGCAGTGTGAGTCTCAACCTAACCTTTCATAGT 1701
1630 ATCAGTATTCAACAACAGTTTGCCTGTTGCGCAGTGTGAGTCTCAACCTAACCTTTCATAGT 1689
1702 AATTCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTTACCATTAACCAATTCGAGTA 1761
1690 AATTCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTTACCATTAACCAATTCGAGTA 1749
1762 CAGTCTACCTCGAAGCATCTACGATGTGAGTCTTCTAGTAAAGTAAACAAACCGATCCCC 1821
1750 CAGTCTACCTCGAAGCATCTACGATGTGAGTCTTCTAGTAAAGTAAACAAACCGATCCCC 1809
1822 CGCAGTGAATCTCTGCTCCAGCCGCTGATGAATGGCAATCCAAAGCTGAACTCCAGCGTG 1881
1810 CGCAGTGAATCTCTGCTCCAGCCGCTGATGAATGGCAATCCAAAGCTGAACTCCAGCGTG 1869
1882 CTGCTGCCCTATGGCGCCGAGTCTCTGAGGACTCTGACGAGGAGTCAAAGGGGCTGGGC 1941
1870 CTGCTGCCCTATGGCGCCGAGTCTCTGAGGACTCTGACGAGGAGTCAAAGGGGCTGGGC 1929
1942 AAGGAGATGGGATGGTACGATGTGAGTCTCCACTCTCCGCGCCCAAGTCCCGAAGT 2001
1930 AAGGAGATGGGATGGTACGATGTGAGTCTCCACTCTCCGCGCCCAAGTCCCGAAGT 1989
2002 GAGGAGCCCACTCCGACGAGCTTCAAGAACCCATGACCTTAAACGGTGTCTAATAGTGCA 2061
1990 GAGGAGCCCACTCCGACGAGCTTCAAGAACCCATGACCTTAAACGGTGTCTAATAGTGCA 2049
2062 GACAGGACAGTGACCCGAAAGAAAACGGCTAGCGCTGTGATGGTCCAGCTGCCAAGGC 2121
2050 GACAGGACAGTGACCCGAAAGAAAACGGCTAGCGCTGTGATGGTCCAGCTGCCAAGGC 2109
2122 CAGCTCTCCCTGCACCTCAGAAATCCCTTGTCTAGGCAAGCGTCTTCTCGAAAGTTG 2181
2110 CAGCTCTCCCTGCACCTCAGAAATCCCTTGTCTAGGCAAGCGTCTTCTCGAAAGTTG 2169
2182 ATGCTCTCTCTTGTCTCTCTCCAGAAACAAATCTTAGAGACCTTCAAGGCTTAGC 2241
2170 ATGCTCTCTCTTGTCTCTCTCCAGAAACAAATCTTAGAGACCTTCAAGGCTTAGC 2229
2242 AACAACTGAAAGGCTGACCGATGAATGATGCACTTGGAGCAGAGAGGGCCCTCC 2301
2230 AACAACTGAAAGGCTGACCGATGAATGATGCACTTGGAGCAGAGAGGGCCCTCC 2289
2302 GAGGACCGGACCGCGGCTCAGCTGCGCGCCGCGCGCGCAATCCCTGAGGAGC 2359
2290 GAGGACCGGACCGCGGCTCAGCTGCGCGCCGCGCGCGCAATCCCTGAGGAGC 2347

RESULT 4

AX882263

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

AX882263
Sequence 17168 from Patent EP1074617.AX882263
AX882263.1 GI:40037087

Homo sapiens (human)

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.

linear

PAT 17-DEB-2003

TITLE	Primers for synthesising full-length cDNA and their use
JOURNAL	Patent: EP 1074617-A 17168 07-FEB-2001; Research Association for Biotechnology (JP)
FEATURES	Location/Qualifiers
source	1..3660
	/organism="Homo sapiens"
	/mol_type="unassigned DNA"
	/db_xref="taxon:9606"
CDS	68..53658
	/note="unnamed protein product"
	/codon_start=1
	/protein_id="CAE92364.1"
	/db_xref="GI:40037088"
	/translation="MTIVDKASSSDPSAYQNGPSSEAYSPGMDAGSASWGAVSL NDVSNHTLSLGPVGAIVGVSSSPDKSPQDQALGDGIAPPKQVLPFSPKICLK WQTHRLGAGLNGNTFANAALQCLTYTPLANYMLSHEHSKTHAEFGCMWCTMQ AHTQALSNPDVTKPMFVNMERIRARHEFNOEDAEHFLQYTVDMQAKCLNLSQ KLDRHTQATLVCOIFGGLRSVKCLNCKGVSDTFDPYLDITLEIKAAQSVNKALQ FVKEQLDGENSKCKSKMWPKASRFTIHRSSNVLTLSLKRFANFTGGKIAKDVKY PEYLDIRPYMSQNGEPIVVLVAVLVHTGFCHAGHYFCYIKASNGLWYQMDISIVS TSDIRSVLSQQAIVFLIRSHDVKNKGELTHPSPQSPRPVVISORVTVNKQAAPG FISQVPEHPMKQKQITISHNKLIVRQCQSPNLHNSLENTKPVPSSTITNSAVQST RSVPIPEHPKQKQITISHNKLIVRQCQSPNLHNSLENTKPVPSSTITNSAVQST SNASTSVSKVTKPIPRSSCSQPVNMKGSKLNSVLVPGAESDSEESKGLGK ENGIGTIVSSHSPQDAEDBEATPHELQEPITLLNGANSADSDPKENGLAPDQSCQ GQALHSENFPAXANGLPGLMPALISLPEDKILETFLRLSNKLGSTDEMSAPGER GPEDRDAGEQSPAESLEEDPAAGLSLSTKKAAPPDPDPGPATKEGAWEMAVAP EEPPSDAEIVGDTAPDLDCPSGLTGDASPLSQDAKGMIAEGPRDSALAEAPGLS PAPPARSEECDEOPLLVHPGSDHARDQDPSQSLGAPAEARPPAPVLDWAPAGHPG DAPSPQERVEDAAPKAPGPSFAKEIGSLRKVDGRYRSTRERSSGSPAPBSRSK TEGHRHRRRTCTPERDQRDRHAPHPHGHGDRLSGERRSLGRCSHHHSHRHSGBEL DWYRHVHTEGERGKREKFPDRPWDRCRYHDRVLYAARDWKPHGGRHERAGL HERPHKDNHRRRCCEPARERHRPSSPRAGAPHALAPHPDRPSHPDRALVAGDNCN LSDRFHEHNGKSKRRHSDSVENSDDSHVEKKARRSEOKDPLEEPKAKGKKK"
ORIGIN	
Query Match	93.9%; Score 2265; DB 6; Length 3660;
Best Local Similarity	99.9%; Pred. No. 0;
Matches 2365; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
QY	6 GTGTGCGCGCGCGCGCGCGCGGAGTGGAGCGAGCGCCGAGCGGTTCAGAG 65
Db	1 GTGTGCGCGCGCGCGCGCGCGGAGTGGAGCGAGCGCGCGGTTCAGAG 60
QY	66 TTGAACAATGACCAATAGTTGACAAAGCTTCTGAATCTTCAGACCATCAGCTTACGAA 125
Db	61 TTGAACAATGACCAATAGTTGACAAAGCTTCTGAATCTTCAGACCATCAGCTTACGAA 120
QY	126 TCAGCCTGGCAGCTCCGAGGCGAGTCTCACCTGGGAGACATGGATGCGAGGTTCTGCCAGCTG 185
Db	121 TCAGCCTGGCAGCTCCGAGGCGAGTCTCACCTGGGAGACATGGATGCGAGGTTCTGCCAGCTG 180
QY	186 GGGTGTGTGTCTTCATTGAATGATGTGCAAAATCACACACTTTCTTTTAGGACCGAGTACC 245
Db	181 GGGTGTGTGTCTTCATTGAATGATGTGCAAAATCACACACTTTCTTTTAGGACCGAGTACC 240
QY	246 TGGTGTGTGTCTTTATTCGAGTTTATCTGTACCTGATTAATCAAAACCATCACCACAAAA 305
Db	241 TGGTGTGTGTCTTTATTCGAGTTTATCTGTACCTGATTAATCAAAACCATCACCACAAAA 300
QY	306 GGATCAAGCCCTAGGTGATGCGATCGCTCCCTCCACAGAAAGTCTTTTCCCATCTGAGAA 365
Db	301 GGATCAAGCCCTAGGTGATGCGATCGCTCCCTCCACAGAAAGTCTTTTCCCATCTGAGAA 360
QY	366 GATTGTCTTAAGTGGCAACAAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAA 425
Db	361 GATTGTCTTAAGTGGCAACAAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAA 420
QY	426 TACCTGTTTTCGAATGAGCACTGCGAGTGTAACTTACACACACCTCTTGGCAATTA 485
Db	421 TACCTGTTTTCGAATGAGCACTGCGAGTGTAACTTACACACACCTCTTGGCAATTA 480
QY	486 CATGCTATCACATGAACACTCCAAAACATGTGATGCAAGAGGCTTTTGTATGATGTGTAC 545
Db	481 CATGCTATCACATGAACACTCCAAAACATGTGATGCAAGAGGCTTTTGTATGATGTGTAC 540
QY	546 AATGCAAGCACATATTACCCAGGCACTCAGTAATCTCTGGGAGCGTTATTTAAACCAATGTT 605
Db	541 AATGCAAGCACATATTACCCAGGCACTCAGTAATCTCTGGGAGCGTTATTTAAACCAATGTT 600
QY	606 TGTTCATCAATGAGATGCGGGGTATAGCTAGGCACTCCGCTTTTGGAAACCAAGAAGATGC 665
Db	601 TGTTCATCAATGAGATGCGGGGTATAGCTAGGCACTCCGCTTTTGGAAACCAAGAAGATGC 660
QY	666 CCATGAATCTCTTCAATACACTGTGTGATGCTATGCAAGGCAATGCTTGAATGCGAGCAA 725
Db	661 CCATGAATCTCTTCAATACACTGTGTGATGCTATGCAAGGCAATGCTTGAATGCGAGCAA 720
QY	726 TAAATTAGACAGACACACCCAGGCCACACTCTGTGTTTGTTCAGATATTTGGAGGATACCT 785
Db	721 TAAATTAGACAGACACACCCAGGCCACACTCTGTGTTTGTTCAGATATTTGGAGGATACCT 780
QY	786 AAGATCTAGAGTCAATATGTTTAAATTCGAAGGCGTTTCAGATACTTTTGATCCATATCT 845
Db	781 AAGATCTAGAGTCAATATGTTTAAATTCGAAGGCGTTTCAGATACTTTTGATCCATATCT 840
QY	846 TGATATAACATTGGAGATAAAGGCTGCTCAGAGTGTCAAAGGCATTGGAGCAGTTTGT 905
Db	841 TGATATAACATTGGAGATAAAGGCTGCTCAGAGTGTCAAAGGCATTGGAGCAGTTTGT 900
QY	906 GAAGCCGGAACAGCTTGTATGGAGAAACTCGTCAAGTCGAGCAAGTGTAAAAAGATGGT 965
Db	901 GAAGCCGGAACAGCTTGTATGGAGAAACTCGTCAAGTCGAGCAAGTGTAAAAAGATGGT 960
QY	966 TCCAGCTTCAAAAGAGGTTCACTATCCATAGATCTCTAAATGTTCTTACACTTTCTCTGAA 1025
Db	961 TCCAGCTTCAAAAGAGGTTCACTATCCATAGATCTCTAAATGTTCTTACACTTTCTCTGAA 1020
QY	1026 ACCTTTTGCATAATTTTACCCTGGGAAAAATTTGCTAAGGATGTGAAATACCTCGATATCT 1085
Db	1021 ACCTTTTGCATAATTTTACCCTGGGAAAAATTTGCTAAGGATGTGAAATACCTCGATATCT 1080
QY	1086 TGATATTGGCCCATATATGTTCTCAACCAAGGAGAGCCATTCCTACGCTCTGTGTATGC 1145
Db	1081 TGATATTGGCCCATATATGTTCTCAACCAAGGAGAGCCATTCCTACGCTCTGTGTATGC 1140
QY	1146 AGTCTCGTCCACACTGCTGTTTAAATTCGCATGCTGGCCATTACTTCTGTACATAAAAGC 1205
Db	1141 AGTCTCGTCCACACTGCTGTTTAAATTCGCATGCTGGCCATTACTTCTGTACATAAAAGC 1200
QY	1206 TAGCAATGGGCTCTGTATCAAATGAATGACTCCATTTGATCTACCGAGTGTATTAGATC 1265
Db	1201 TAGCAATGGGCTCTGTATCAAATGAATGACTCCATTTGATCTACCGAGTGTATTAGATC 1260
QY	1266 GGTACTCAGCCCAACAGCCATATGCTCTTTTATATCAGGTCCTCATGATGTGAAAAATGG 1325
Db	1261 GGTACTCAGCCCAACAGCCATATGCTCTTTTATATCAGGTCCTCATGATGTGAAAAATGG 1320
QY	1326 AGGTGAATCTACTCATCCCACTAGCCCGGCGAGTCTCTCCCGCCCGCCGTCATCAG 1385
Db	1321 AGGTGAATCTACTCATCCCACTAGCCCGGCGAGTCTCTCCCGCCCGCCGTCATCAG 1380
QY	1386 TCAGCGGTTGTGTCACCAACAAAGGCTGCGCCAGGCTTTTATCGGACCAAGCTTCCCTC 1445
Db	1381 TCAGCGGTTGTGTCACCAACAAAGGCTGCGCCAGGCTTTTATCGGACCAAGCTTCCCTC 1440
QY	1446 TCACATGATAAAGAAATCCACCTCACTTAAATGGGACTGGACCATTTGAAGACACGCCAAG 1505
Db	1441 TCACATGATAAAGAAATCCACCTCACTTAAATGGGACTGGACCATTTGAAGACACGCCAAG 1500
QY	1506 CAGTTCCATGTGCGAGTCTTAAACGGGAAATTCAGTGTCAACAGGCGTAGTCTCTGTTAATGC 1565
Db	1501 CAGTTCCATGTGCGAGTCTTAAACGGGAAATTCAGTGTCAACAGGCGTAGTCTCTGTTAATGC 1560
QY	1566 TTGAGCTTCTGTCCAAACTGGTCAAGTTAATAGTCTCTCAGTGATCCCGAGACATCTCTAA 1625

```
Db 1561 TTCAGCTTCTGTCCAAATGGTTCAGTAAATAGGTCCTCAGTGATCCAGAAATCCTAA 1620
QY 1626 GAAACAAAATTAACATCAGTATTCACACAGTGTGCTGTTCCGCCAGTGTGAGTCTCA 1685
Db 1621 GAAACAAAATTAACATCAGTATTCACACAGTGTGCTGTTCCGCCAGTGTGAGTCTCA 1680
QY 1686 ACCTAACCTTTCATAGTAATTTCTTTGGAGAACCCCTACCAAGCCGCTTCCTCTTCTACCAT 1745
Db 1681 ACCTAACCTTTCATAGTAATTTCTTTGGAGAACCCCTACCAAGCCGCTTCCTCTTCTACCAT 1740
QY 1746 TACCAATTTGCGAGTACAGTCTACCTCGAACCGCATCTACGATGTCAGTGTCTAGTAAAGT 1805
Db 1741 TACCAATTTGCGAGTACAGTCTACCTCGAACCGCATCTACGATGTCAGTGTCTAGTAAAGT 1800
QY 1806 AACAAAACCGATCCCCCGCAGTGAATCTGCTCCCGAGCCGCTGATGATGCAATCCAA 1865
Db 1801 AACAAAACCGATCCCCCGCAGTGAATCTGCTCCCGAGCCGCTGATGATGCAATCCAA 1860
QY 1866 GCTGAATCCAGCGTGTGGTGCCTATGCGCGCGAGTCTCTGAGGACTCTGACGAGGA 1925
Db 1861 GCTGAATCCAGCGTGTGGTGCCTATGCGCGCGAGTCTCTGAGGACTCTGACGAGGA 1920
QY 1926 GTCAAGGGGCTGGCAAGAGATGGATTTGTAGTATGATGTCAGTCTCCACTCTCCCG 1985
Db 1921 GTCAAGGGGCTGGCAAGAGATGGATTTGTAGTATGATGTCAGTCTCCACTCTCCCG 1980
QY 1986 CCAAGATGCCGAAGATGAGAGGCCACTCGCGCAGGCTTCAAGAACCCATGACCTTAA 2045
Db 1981 CCAAGATGCCGAAGATGAGAGGCCACTCGCGCAGGCTTCAAGAACCCATGACCTTAA 2040
QY 2046 CGGTGCTAATAGTGCAGACGACAGTGCACCGCAAGAAACGGCTAGCGCTGATGG 2105
Db 2041 CGGTGCTAATAGTGCAGACGACAGTGCACCGCAAGAAACGGCTAGCGCTGATGG 2100
QY 2106 TGCAGCTGCCAAGGCGAGCTGCCCTGCACTCAGAAAATCCCTTGTAAAGGCAACGG 2165
Db 2101 TGCAGCTGCCAAGGCGAGCTGCCCTGCACTCAGAAAATCCCTTGTAAAGGCAACGG 2160
QY 2166 TCTTCTGGAAGTGTATGCTGCTCTTGTCTCTCTCCAGAGACAAATCTTAGA 2225
Db 2161 TCTTCTGGAAGTGTATGCTGCTCTTGTCTCTCTCCAGAGACAAATCTTAGA 2220
QY 2226 GACCTTCAGGCTTAGCAACAACTGAAAGGCTGACGGATGAAATGAGTGCACCTGGAGC 2285
Db 2221 GACCTTCAGGCTTAGCAACAACTGAAAGGCTGACGGATGAAATGAGTGCACCTGGAGC 2280
QY 2286 AGAGAGGGGCTTCCGAGGACGCGACGCGAGCTCAGCCTGGCAGCCCGCCGCGGA 2345
Db 2281 AGAGAGGGGCTTCCGAGGACGCGACGCGAGCTCAGCCTGGCAGCCCGCCGCGGA 2340
QY 2346 ATCCCTGGAGGAGCCAGATGCGGCGCG 2372
Db 2341 ATCCCTGGAGGAGCCAGATGCGGCGCG 2367

RESULT 5
BD159625
LOCUS 3660 bp DNA linear PAT 17-JAN-2003
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD159625
VERSION BD159625.1 GI:27865383
KEYWORDS JP 2002191363-A/14468.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3660)
Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 14468 09-JUL-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
```

```
PN JP 2002191363-A/14468
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
PI SAITO,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUO OHSUKI
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10,
PC C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof PH key
FT CDS (68)..(3658).
FEATURES
    source
        1..3660
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
ORIGIN
Query Match 93.9%; Score 2265; DB 6; Length 3660;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2365; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 6 GTGTGCGGCGCGCGCGCGCGCGCGGAGGAGTGAGCGCGCGCGCGCGCGCGGTTCAGAG 65
Db 1 GTGTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTTCAGAG 60
QY 66 TTGAACAATGACCATAGTTGACAAAGCTTCTGAACTTTCAGAGCCCATCAGCCTATCAGAA 125
Db 61 TTGAACAATGACCATAGTTGACAAAGCTTCTGAACTTTCAGAGCCCATCAGCCTATCAGAA 120
QY 126 TCAGCCTGCGAGCTCCGAGGCGAGTCTCAGTGCAGACATGATGAGGCTTCTGCCAGCTG 185
Db 121 TCAGCCTGCGAGCTCCGAGGCGAGTCTCAGTGCAGACATGATGAGGCTTCTGCCAGCTG 180
QY 186 GGTGCTGTGCTTCAATGAATGAATGTCAAATCAACACACTTCTTTAGGACCACTACC 245
Db 181 GGTGCTGTGCTTCAATGAATGAATGTCAAATCAACACACTTCTTTAGGACCACTACC 240
QY 246 TGTGCTGTAGTATTTATTCAGGTTTCATCTGATCTGATTAATCAAAACCATCACCACAAA 305
Db 241 TGTGCTGTAGTATTTATTCAGGTTTCATCTGATCTGATTAATCAAAACCATCACCACAAA 300
QY 306 GGATCAAGCCCTAGGTGATGGCATCGCTCTCCACAGAAAGTCTTTTCCCATCTCAGAA 365
Db 301 GGATCAAGCCCTAGGTGATGGCATCGCTCTCCACAGAAAGTCTTTTCCCATCTCAGAA 360
QY 366 GATTGTCTTAAGTGGCAACAACTCATAGATGGAGCTGGGCTCCAGAAATTTGGGCAA 425
Db 361 GATTGTCTTAAGTGGCAACAACTCATAGATGGAGCTGGGCTCCAGAAATTTGGGCAA 420
QY 426 TACCTGTTTGGCAATGCGAGCACTCGAGTGTAACTTACCTACACACACCTCTTCCCAATTA 485
Db 421 TACCTGTTTGGCAATGCGAGCACTCGAGTGTAACTTACCTACACACACCTCTTCCCAATTA 480
QY 486 CATGCTATCAGATGAACACTCCAAAACATGTATGAGAGGCTTTTGTATGATGTGTAC 545
Db 481 CATGCTATCAGATGAACACTCCAAAACATGTATGAGAGGCTTTTGTATGATGTGTAC 540
QY 546 AATGCAAGCATATTTACCGAGCACTCAGTAATCTCTGGGAGCGTTATTAAACCAATGTT 605
Db 541 AATGCAAGCATATTTACCGAGCACTCAGTAATCTCTGGGAGCGTTATTAAACCAATGTT 600
QY 606 TGTCAATCAATGAGTGGCGGCTATAGCTAGGACCTCCGTTTGGAAAACCAAGAAGATGC 665
Db 601 TGTCAATCAATGAGTGGCGGCTATAGCTAGGACCTCCGTTTGGAAAACCAAGAAGATGC 660
QY 666 CCATGAATTCCTTCAATACACTGTTGATGCTATGAGAGAGCATCTTGAATGGCAGCAA 725
Db 661 CCATGAATTCCTTCAATACACTGTTGATGCTATGAGAGAGCATCTTGAATGGCAGCAA 720
```

QY 726 TAAATTAGACAGACACACCCAGGCCACCACTCTTGTGTGTCAGATATTTTGGAGGATACCT 785
DB 721 TAAATTAGACAGACACACCCAGGCCACCACTCTTGTGTGTCAGATATTTTGGAGGATACCT 780
QY 786 AAGATCTAGAGTCAAAATGTTTAAATTCGAAGGGCGTTTCAGATACCTTTTGCATCATCT 845
DB 781 AAGATCTAGAGTCAAAATGTTTAAATTCGAAGGGCGTTTCAGATACCTTTTGCATCATCT 840
QY 846 TGATATAACATTTGGAGATAAAGGCTGCTCAGAGTGTCAAAAGGCATTTGGAGGAGTTTGT 905
DB 841 TGATATAACATTTGGAGATAAAGGCTGCTCAGAGTGTCAAAAGGCATTTGGAGGAGTTTGT 900
QY 906 GAAGCCGGAACAGCTTATGAGGAGAACTCGTACAGTGCAGCAAGTGAAGAGATGGT 965
DB 901 GAAGCCGGAACAGCTTATGAGGAGAACTCGTACAGTGCAGCAAGTGAAGAGATGGT 960
QY 966 TCCAGCTTCAAGAGAGTTCACTATCCATAGATCTCTAAATGTTCTTACACTTTCTCTGAA 1025
DB 961 TCCAGCTTCAAGAGAGTTCACTATCCATAGATCTCTAAATGTTCTTACACTTTCTCTGAA 1020
QY 1026 ACGTTTTCGCAAAATTTACCGGTGGAATAATGCTAAGAGTGAATAACCTGAGTATCT 1085
DB 1021 ACGTTTTCGCAAAATTTACCGGTGGAATAATGCTAAGAGTGAATAACCTGAGTATCT 1080
QY 1086 TGATATTCGGCCATATATGCTCAACCAACGAGAGCCATTCGTACGTCCTTGTATGC 1145
DB 1081 TGATATTCGGCCATATATGCTCAACCAACGAGAGCCATTCGTACGTCCTTGTATGC 1140
QY 1146 AGTCTGCTCCACACTGTTTAAATTTGTCATGCTGGCCATTTACTTCTGCTACATAAAGC 1205
DB 1141 AGTCTGCTCCACACTGTTTAAATTTGTCATGCTGGCCATTTACTTCTGCTACATAAAGC 1200
QY 1206 TAGCAATGGCTCTGGTATCAAAATGAATGATCTCAATTTGTATCTACAGTATTTAGATC 1265
DB 1201 TAGCAATGGCTCTGGTATCAAAATGAATGATCTCAATTTGTATCTACAGTATTTAGATC 1260
QY 1266 GGTACTCAGCCAAACAGGCTATGCTCTTTATATCAGTCCCATGATGTGAATAATGG 1325
DB 1261 GGTACTCAGCCAAACAGGCTATGCTCTTTATATCAGTCCCATGATGTGAATAATGG 1320
QY 1326 AGGTGAATCTACTCATCCACCCATAGCCCGGCGAGTCTCTCCCGCCCGCTCATCAG 1385
DB 1321 AGGTGAATCTACTCATCCACCCATAGCCCGGCGAGTCTCTCCCGCCCGCTCATCAG 1380
QY 1386 TCAGCGGTTGTACCAACAAACAGGCTGGCGAGGCTTTATCGACACACAGCTTCCCTC 1445
DB 1381 TCAGCGGTTGTACCAACAAACAGGCTGGCGAGGCTTTATCGGACACACAGCTTCCCTC 1440
QY 1446 TCACATGATAAGAAATCCACCTCACTTAAATGGGACTGGACCATTTGAAGACACGCCAAG 1505
DB 1441 TCACATGATAAGAAATCCACCTCACTTAAATGGGACTGGACCATTTGAAGACACGCCAAG 1500
QY 1506 CAGTTTCCATGTGAGTTCCTAAGCGGAATTTCCAGTGTCAACAGGGCTAGTCTGTTTAAATGC 1565
DB 1501 CAGTTTCCATGTGAGTTCCTAAGCGGAATTTCCAGTGTCAACAGGGCTAGTCTGTTTAAATGC 1560
QY 1566 TTCAGCTTCTGTCOAAAACAGTGTGAGTAAATAGTCTCCTCAGTATCCAGACACATCTTAA 1625
DB 1561 TTCAGCTTCTGTCOAAAACAGTGTGAGTAAATAGTCTCCTCAGTATCCAGACACATCTTAA 1620
QY 1626 GAAACAAAATTTACATCAGTATTTACAAACAGTTGCTGTTTCGCCAGTGTGAGTCTCA 1685
DB 1621 GAAACAAAATTTACATCAGTATTTACAAACAGTTGCTGTTTCGCCAGTGTGAGTCTCA 1680
QY 1686 ACCTAACCTTCAATAGTAAATTTCTTTGGAGAACCTTACCAAGCCCGTTCCTCTTCTACCAT 1745
DB 1681 ACCTAACCTTCAATAGTAAATTTCTTTGGAGAACCTTACCAAGCCCGTTCCTCTTCTACCAT 1740
QY 1746 TACCAATTTGAGTACAGTCTACCTCGAACCATCTACAGTGTGAGTCTTCTAGTAAAGT 1805
DB 1741 TACCAATTTGAGTACAGTCTACCTCGAACCATCTACAGTGTGAGTCTTCTAGTAAAGT 1800
QY 1806 AACAAAAACGATCCCCCGCAGTGAATCTCTGCTCCAGCCCGTGTGATGAATGGCAAAATCCAA 1865

DB 1801 AACAAAAACGATCCCCCGCAGTGAATCTCTGCTCCAGCCCGTGAATGAATGCAATCCAA 1860
QY 1866 GCTGAATCTCAGGCTGCTGCTGCTATGCGCGGAGTCTCTGAGGACTCTGACGAGGA 1925
DB 1861 GCTGAATCTCAGGCTGCTGCTGCTATGCGCGGAGTCTCTGAGGACTCTGACGAGGA 1920
QY 1926 GTCAAGGGGCTGGGCAAGGAGAAATGGGATTTGATTCGATTTGAGCTCTCCCGG 1985
DB 1921 GTCAAGGGGCTGGGCAAGGAGAAATGGGATTTGATTCGATTTGAGCTCTCCCGG 1980
QY 1986 CCAAGATGCCAAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCTTAAA 2045
DB 1981 CCAAGATGCCAAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCTTAAA 2040
QY 2046 CGGTGCTTAATGTCAGACAGGACGACGACGACGACGACGACGACGACGACGACGACG 2105
DB 2041 CGGTGCTTAATGTCAGACAGGACGACGACGACGACGACGACGACGACGACGACGACG 2100
QY 2106 TGCAGCTGCCAAGGCCAGGCTGCCCTGCACTCAGAAAAATCCCTTTGCTAAGGCAACGG 2165
DB 2101 TGCAGCTGCCAAGGCCAGGCTGCCCTGCACTCAGAAAAATCCCTTTGCTAAGGCAACGG 2160
QY 2166 TCTTCTCGAAAAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2225
DB 2161 TCTTCTCGAAAAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2220
QY 2226 GACCTTCAGGCTTACCAACAACTGAAAGGCTCCAGCGATGAATGAGTGCACCTGGAGC 2285
DB 2221 GACCTTCAGGCTTACCAACAACTGAAAGGCTCCAGCGATGAATGAGTGCACCTGGAGC 2280
QY 2286 AGAGAGGGGCTCTCCGAGGACCGGACGCGAGCTCAGCTGCGAGCCCGCGCGCGA 2345
DB 2281 AGAGAGGGGCTCTCCGAGGACCGGAGCGCGAGCTCAGCTGCGAGCCCGCGCGCGA 2340
QY 2346 ATCCCTGGAGGAGCCAGATGCGGCGC 2372
DB 2341 ATCCCTGGAGGAGCCAGATGCGGCGC 2367

RESULT 6

AK022759 3660 bp mRNA linear PRI 01-AUG-2002
LOCUS Homo sapiens cDNA FLJ12697 fis, clone NT2RP1000522, weakly similar
to UBQJUITIN CARBOXYL-TERMINAL HYDROLASE DUB-1 (EC 3.1.2.15).

AK022759.1 GI:10434351

VERSION oligo capping; fis (full insert sequence).

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H.,
Wagatsuma, M., Hosokiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K.,
Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,
Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,
Nakamura, Y., Nagahari, K., Masuno, Y., Ninomiya, K. and Iwayanagi, T.
NEDO human cDNA sequencing project

Unpublished

2 (bases 1 to 3660)

Isogai, T. and Otsuki, T.

Direct Submission

Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,

Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of

International Trade and Industry of Japan; cDNA full insert

sequencing; Research Association for Biotechnology; cDNA library

construction, 5'- & 3'-end one pass sequencing and clone selection;

Helix Research Institute (supported by Japan Key Technology Center

etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.

FEATURES

source Location/Qualifiers

1..3660

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="NT2RP1000522"

/cell_line="NT2"

/cell_type="teratocarcinoma"

/clone_lib="NT2RP1"

/notes="cloning vector: pUC19FL3-mRNA from NT2 neuronal precursor cells after 48-hours retinoic acid (RA) induction."

68..>3660

/note="unnamed protein product"

/codon_start=1

/protein_id="BAB14232.1"

/db_xref="GI:10434352"

/translation="MTIVDKASESDPSAVONOPGSSEAVPCGMDAGSAGVSSLL
NDVSNHTLSLGPVPGAVVSSSSVPDKSPKQDQALGDGIAPPQKVLPPSEKI CLK
WOOTHEVAGLONLNTCFANAALQCLTTPPLANTMLSEHSKTCCHAGFCNMCTMQ
AHTQALSNPQVVKMFVNEMRRIARHFRQNEDEAHEFLQYTVDMQKCLNSN
KIDRHTQATLVCOIFGGVLRVKNCKNGSDTFDYLDTLETKAAQSNKALEQ
FVKPEQLDGENSYKCKSKMVPASXRFTHRSSNVLTLILKRFANDTGTGKIAKDVY
PYLDIRPYMSQPNQGEPIVVVLVAVLVHTGFENCHAGHYCYIKASNLGLYQWDSIVS
TSDIRSLVSOQAVLFIIRSHDVKNQSGELTHPTSPGQSSPRPVISQRVVTNKQAAVG
FIGPOLPSHMIKNPPLHNGTGPLKDTPPSSMSPNNGSNVNRAPVNASVONMSVN
RSVIPLEHPKMKITISINHKLPVROCOQPNLHNSLENPTKPVPSSTITNSAVGT
SNASTSVSKTKITISINHKLPVROCOQPNLHNSLENPTKPVPSSTITNSAVGT
ENGI GTI VSHSGPQDAEAEATPHELOPPTTLINGANSADSDPKENGLAPDGASCO
GQPALHSENFPAKGLPKLAPLISLPEDKILFTFRLSNLKGSTDEMSAPGAE
GPEPDRDABPQGSAAESLEPDAAGLSSTKAPPPDPDPTATKEGAWEMAVP
REPPSAGDI VGDTPAPLDPGSLTGDASLSLQDAKMLAEPPDLSALAEPEGLS
PAPPARSEPCQPLIPLSPGSDHDAQDPSQSLGAPEAAERPPAPVLDMAPAHPEG
DAEPSGERVEDAAKPAKPGYPAKEKIGSLRKVDGRHYRERRSSSGEPARESRK
TGHRRHRTCTPREDRDQRAHEPHHGHGDRLSFGERRSLGRCSHHHSRHSVGL
DWRHHYTGEGRGREKREYRPPRWRDRCRYVDRYVAARDWKPFGHGRHERAGL
HERPHKDNHGRGCEPAPERREHPPSPRAGAPHALAHPDRFSDHRTALVAGDNCN
LSDRFEHENGKSRKRRHDSVENSISHVEKKARRSEQKDLPEPKAKKHKS"

ORIGIN

Query Match 93.9%; Score 2265; DB 9; Length 3660;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2365; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	426	TACCTGTTTTGCCAATGCAGCACTGCAGTGTGTTAACTTACACACACCTCTTGCCAATTA	485
DB	421	TACCTGTTTTGCCAATGCAGCACTGCAGTGTGTTAACTTACACACACCTCTTGCCAATTA	480
QY	486	CATGCTATCATGAGAACACCTCCAAAACATGTCATGCAGAGGCTTTTGTATGATGTGTAC	545
DB	481	CATGCTATCATGAGAACACCTCCAAAACATGTCATGCAGAGGCTTTTGTATGATGTGTAC	540
QY	546	AATGCAAGACATATATACCCAGGCACCTCAGTAATCTCTGGGGAGCGTTATTAACCAATGTT	605
DB	541	AATGCAAGACATATATACCCAGGCACCTCAGTAATCTCTGGGGAGCGTTATTAACCAATGTT	600
QY	606	TGTCATCAATGAGATCGGGGTATAGCTAGGCACCTCCGCTTTTGGAAACCAAGAGATGC	665
DB	601	TGTCATCAATGAGATCGGGGTATAGCTAGGCACCTCCGCTTTTGGAAACCAAGAGATGC	660
QY	666	CCATGATTCCTTCAATACACTGTTGATGCTATGCGAAGAGCATGCTTGAATGGCAGCAA	725
DB	661	CCATGATTCCTTCAATACACTGTTGATGCTATGCGAAGAGCATGCTTGAATGGCAGCAA	720
QY	726	TAAATTTAGACAGACACACCCAGGCCACCACTCTCTGTTGTCAGATATTTGGAGGATACCT	795
DB	721	TAAATTTAGACAGACACACCCAGGCCACCACTCTCTGTTGTCAGATATTTGGAGGATACCT	780
QY	786	AGATCTAGAGTCAAAATGTTTAAATTTGCAAGGGCGTTTTCAGATCTTTTGCATATCT	845
DB	781	AGATCTAGAGTCAAAATGTTTAAATTTGCAAGGGCGTTTTCAGATCTTTTGCATATCT	840
QY	846	TGATATAACATTTGGAGATAAAGCTCTCAGAGTGTCACCAAGGCATTTGGAGCAGTTGT	905
DB	841	TGATATAACATTTGGAGATAAAGCTCTCAGAGTGTCACCAAGGCATTTGGAGCAGTTGT	900
QY	906	GAAGCGGAAACAGCTTGATGGGAAAATCTGTCAGAGTGCAGCAAGTGTAAAGATGCT	965
DB	901	GAAGCGGAAACAGCTTGATGGGAAAATCTGTCAGAGTGCAGCAAGTGTAAAGATGCT	960
QY	966	TCCAGCTTCAAAAGAGGTTCACTATCCATAGATCTCTAATGTTCTTACACTTTCTCGAA	1025
DB	961	TCCAGCTTCAAAAGAGGTTCACTATCCATAGATCTCTAATGTTCTTACACTTTCTCGAA	1020
QY	1026	ACGTTTTCGAAATTTTACCGGTGGAAAATTCGTAAGAGTGTGAANAATCCCTGAGTATCT	1085
DB	1021	ACGTTTTCGAAATTTTACCGGTGGAAAATTCGTAAGAGTGTGAANAATCCCTGAGTATCT	1080
QY	1086	TGATATTCGGCCCATATATGTCCTCAACCCAGGAGGCAATGTCAGTCTTGTATGC	1145
DB	1081	TGATATTCGGCCCATATATGTCCTCAACCCAGGAGGCAATGTCAGTCTTGTATGC	1140
QY	1146	AGTGTCTGTCACACTGCTGTTTAAATTTGCCATGCTGGCCCATTACTTCTGCTACATAAAGC	1205
DB	1141	AGTGTCTGTCACACTGCTGTTTAAATTTGCCATGCTGGCCCATTACTTCTGCTACATAAAGC	1200
QY	1206	TAGCAATGCGCTCTGTTATCAAAATGATCTCCATTTGATCTACAGTGATATTAGATC	1265
DB	1201	TAGCAATGCGCTCTGTTATCAAAATGATCTCCATTTGATCTACAGTGATATTAGATC	1260
QY	1266	GGTACTCAGCCCAACAGGCTATGTCCTCTTTTATATCAGGTCCCATGATGTGAAAATGG	1325
DB	1261	GGTACTCAGCCCAACAGGCTATGTCCTCTTTTATATCAGGTCCCATGATGTGAAAATGG	1320
QY	1326	AGGTGAATTTACTCATCCCAATAGCCCGGCGAGTCTCTCCCGCCCGCTCATCAG	1385
DB	1321	AGGTGAATTTACTCATCCCAATAGCCCGGCGAGTCTCTCCCGCCCGCTCATCAG	1380
QY	1386	TCAGGGGTGTTGACCAACAAAGGCTGCGCCAGGCTTTATCGGACCAAGCTTCCCTC	1445
DB	1381	TCAGGGGTGTTGACCAACAAAGGCTGCGCCAGGCTTTATCGGACCAAGCTTCCCTC	1440
QY	1446	TCACATGATAAAGAAATCCACCTCACTTAAATGGGACCTGGACATTTGAAAGACAGCCAG	1505
DB	1441	TCACATGATAAAGAAATCCACCTCACTTAAATGGGACCTGGACATTTGAAAGACAGCCAG	1500

QY	1506	CAGTTCATGTCAGTCCCTAAACGGGAATTCAGTGTCAACAGGCTAGTCTCTGTTAATGC	1565
Db	1501	CAGTTCATGTCAGTCCCTAAACGGGAATTCAGTGTCAACAGGCTAGTCTCTGTTAATGC	1560
QY	1566	TTTCAGTCTTCGTCCTAAACCTGTCAGTGTCAATAGTCTCAGTGTCCAGAACATCCTAA	1625
Db	1561	TTTCAGTCTTCGTCCTAAACCTGTCAGTGTCAATAGTCTCAGTGTCCAGAACATCCTAA	1620
QY	1626	GAAACAAAAATACAAATCAGTATTCACAAACAGTTGCTGTTGCCAGTGTCAAGTCTCA	1685
Db	1621	GAAACAAAAATACAAATCAGTATTCACAAACAGTTGCTGTTGCCAGTGTCAAGTCTCA	1680
QY	1686	ACTAACCTTCATAGTAATTCCTTTGGAGAACCTTACCAAGCCGTTCCCTCTCTACCAT	1745
Db	1681	ACTAACCTTCATAGTAATTCCTTTGGAGAACCTTACCAAGCCGTTCCCTCTCTACCAT	1740
QY	1746	TACCAATTCGTCAGTACAGTCTACCTCGAAGCATCTACAGTGTCAAGTCTCTAGTAAAGT	1805
Db	1741	TACCAATTCGTCAGTACAGTCTACCTCGAAGCATCTACAGTGTCAAGTCTCTAGTAAAGT	1800
QY	1806	AACAAACCGATCCCGCGAGTGAATCTCTCCAGCCGTCATGAATGGCAATCCAA	1865
Db	1801	AACAAACCGATCCCGCGAGTGAATCTCTCCAGCCGTCATGAATGGCAATCCAA	1860
QY	1866	GCTGAATCCAGTGTGTCGTCCTATGGCGAGTCTCTGAGACTCTGACGAGGA	1925
Db	1861	GCTGAATCCAGTGTGTCGTCCTATGGCGAGTCTCTGAGACTCTGACGAGGA	1920
QY	1926	GTCAAGGGGCTGGCGAGGAGATGGATTGGTACGATTTGAGCTCCCACTCTCCCGG	1985
Db	1921	GTCAAGGGGCTGGCGAGGAGATGGATTGGTACGATTTGAGCTCCCACTCTCCCGG	1980
QY	1986	CCAAGATGCCAAGATGAGGAGGCCACTCCGACGAGTTCAAGAACCCAGACCTAAA	2045
Db	1981	CCAAGATGCCAAGATGAGGAGGCCACTCCGACGAGTTCAAGAACCCAGACCTAAA	2040
QY	2046	CGGTGCTAATAGTCAGACGACGACGACGACGACGACGACGACGACGACGACGACG	2105
Db	2041	CGGTGCTAATAGTCAGACGACGACGACGACGACGACGACGACGACGACGACGACG	2100
QY	2106	TGCCAGTGTCCCAAGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2165
Db	2101	TGCCAGTGTCCCAAGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2160
QY	2166	TCTTCTCGAAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2225
Db	2161	TCTTCTCGAAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2220
QY	2226	GACCTTCAGGCTTAGCAACAACTGAAAGGCTCGACGATGAAATGAGTGCACCTGGAGC	2285
Db	2221	GACCTTCAGGCTTAGCAACAACTGAAAGGCTCGACGATGAAATGAGTGCACCTGGAGC	2280
QY	2286	AGAGAGGGGCTCCCGAGGACCGGACCGGAGCTTCAGGCTTCAGGCTTCGCGCGCGA	2345
Db	2281	AGAGAGGGGCTCCCGAGGACCGGACCGGAGCTTCAGGCTTCAGGCTTCGCGCGCGA	2340
QY	2346	ATCCTTCGAGGAGCCAGATGCGGCGC	2372
Db	2341	ATCCTTCGAGGAGCCAGATGCGGCGC	2367
RESULT 7	BC060846	3669 bp	PRI 12-NOV-2003
LOCUS	BC060846	3669 bp	PRI 12-NOV-2003
DEFINITION	Homo sapiens ubiquitin specific protease 42, mRNA (cDNA clone IMAGE:30343487), partial cds.		
ACCESSION	BC060846		
VERSION	BC060846.1	GI:38173811	
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Hominidae; Homo.		

REFERENCE	1 (bases 1 to 3659)
AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zebberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, D., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Udwin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., R.A., Willalson, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Bicknell, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE	22388257
PUBMED	12477932
REFERENCE	2 (bases 1 to 3669)
AUTHORS	Strausberg, R.
TITLE	Direct Submission
JOURNAL	Submitted (03-NOV-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: caaphs@email.nih.gov Tissue Procurement: Dr. Stefan Hansson cDNA Library Preparation: Michael Brownstein / Ted Udwin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www-shgc.stanford.edu Contact: (Dickson, Mark) mcd@paxil.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 133 Row: 9 Column: 6
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Genomescan gene prediction, Similarity but not identity to protein.

FEATURES	Location/Qualifiers
source	1..3669 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:30343487" /tissue_type="Placenta, normal" /clone_lib="NIH MGC_147" /lab_host="DH10B" /note="Vector: pBluescript"
gene	1..->3659 /gene="USP42" /note="synonym: FLJ12697" /db_xref="LocusID:84132"
CDS	153..->3669 /gene="USP42" /codon_start=1 /product="USP42 protein" /protein_id="AAH60846.1" /db_xref="GI:38173812" /db_xref="LocusID:84132"

/translation="MDAGSASWGAVSSINDVSNHTLSLGPVPGVAVVYSSSSVDPKSP
SPQDQDGGIAPPQKVPSEKICLKWOOTHEVAGLQNLGNTCFANALQCLTYT
PPLANYMLSEHSKTHAEGFCWCTMQAHTQALSPGVDIKPMFVINEMRRJARHF
REGQEDAEHFQYTVDMQKACLNGSKDHRHTQATILVCQIPGGYILSRVCLNCK
GVSDTFYLDITLIEIKAAQSVNKLQFVKPEQDGENSYKCSKCKMVPASKRFTI
HRSNVLTLSLRFANFTGKIADVKYPIYLDIRPMSQPNGEPIVYVLYAVLVHTG
FNCHAGHYCYIKASNLGWTQMNDISIVSTDIRSVLSQQAIVLPIRSHDVKNGEILT
HPHSPGQSSPRVPIVNSORVWNKQAAPEFTGPQLPSHMIKNPPEHLNGTGPSPSS
MSSPNGSNVNRASPVNASVONWSVRSVTEHPKQKIIYISHNKLPVRQCQSQ
PNLHNSLENPTEKPVSESTINSVQSTSNASTMSVSKVTKEIPRESQSPVMNGK
SKLNSVLYPYGAESSDSDESKGLGKENGIGTIVSSHSPGQDAEDEATPHELOSP
MTUNGANSDDSKENGLAPDQACQGPALHSENPFAKANGLPKMLPAPLLSIP
EDKILFELNSLKGTDMSAPGABERPEPDAEPQPGSAESLEPDAAGLS
STKAPPPRPDPTKGBAWEAMAVAPPEPPPSAGEDIVGDTAPDLCPDPSLTGDA
SPLSQDAKMIASGRDLSALAEAPGLSPAPPSAREPCQPLVHPSPGDHDAQDP
SOSLIGAPEAERPPAPVLDMAPEGHEDAEPSFGERVEDAAAPKAPGPSPAKEKIGS
LRKYDGHYARSKRSESPARESKSTEGHRRHRRRTCPREDRQDRHAPBHPGH
GKRLSPGERRSLGRSHHSRSGVELDWVRHYRTEGERGWREKFYDPRPRWDR
YHDRYALVAARDWKPFHGGREHERAGLHPRKHDRNRGRCCEPARERHRPSR
AGAPHALAPDRFESHDRTALVAGDNCNLSDRFHEHENGSKRKRHDSVENSUSHVK
KARRSQKDPLEBPKAKHKHKKKKK"
393 -1289
misc_feature
/gene="USP42"
/note="UCH; Region: Ubiquitin carboxyl-terminal hydrolase"
/db_xref="CDD:pfam00443"

ORIGIN

	Query Match	93.7%; Score 2259; DB 9; Length 3669;
	Best Local Similarity	100.0%; Pred. No. 0;
	Matches 2309; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY	63	GAGTTGAACAATGACCATAGTGACAAAAGCTTCTGAAATCTTCAGACCCCATCAGCCTATCA 122
DB	53	GAGTTGAACAATGACCATAGTGACAAAAGCTTCTGAAATCTTCAGACCCCATCAGCCTATCA 112
QY	123	GAATCAGCCTGGCAGCTCCGAGGCACTCACTCGGAGACATGGATGCGAGTTCTGCCAG 182
DB	113	GAATCAGCCTGGCAGCTCCGAGGCACTCACTCGGAGACATGGATGCGAGTTCTGCCAG 172
QY	183	CTGGGCTGCTGTCTTCAATGATGTCGAAATGTCGAAATGTCGAAATGTCGAAATGTCGAAAT 242
DB	173	CTGGGCTGCTGTCTTCAATGATGTCGAAATGTCGAAATGTCGAAATGTCGAAATGTCGAAAT 232
QY	243	ACCTGGTGTGTAGTTTATTCGAGTTTCACTGTACCTGTAAATCAAAACCATCACCA 302
DB	233	ACCTGGTGTGTAGTTTATTCGAGTTTCACTGTAAATCAAAACCATCACCA 292
QY	303	AAAGGATCAAGCCCTAGGTGATGGCATGCTCTCCACAGAAAGTTCTTTTCCCATCTGA 362
DB	293	AAAGGATCAAGCCCTAGGTGATGGCATGCTCTCCACAGAAAGTTCTTTTCCCATCTGA 352
QY	363	GAAGATTGTCTTAAGTGGCAACCAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGG 422
DB	353	GAAGATTGTCTTAAGTGGCAACCAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGG 412
QY	423	CAATACCTGTTTGGCAATGACGACTGAGGTGTTTAACTACACACCACTCTTGGCCAA 482
DB	413	CAATACCTGTTTGGCAATGACGACTGAGGTGTTTAACTACACACCACTCTTGGCCAA 472
QY	483	TTAGTGTATCATATGAACACTCCAAACACTGTATGATGAGAGGCTTTTGTATGATGTG 542
DB	473	TTAGTGTATCATATGAACACTCCAAACACTGTATGATGAGAGGCTTTTGTATGATGTG 532
QY	543	TACATGAAGCACATATTAACCGGCACTAGTAATCTCTGGGAGCGTTTATTAACCAAT 602
DB	533	TACATGAAGCACATATTAACCGGCACTAGTAATCTCTGGGAGCGTTTATTAACCAAT 592
QY	603	GTTTGTATCAATGAGATGCGGCGTATAGTAGGCACTCCGTCCTTTTGGAAACCAAGAGA 662
DB	593	GTTTGTATCAATGAGATGCGGCGTATAGTAGGCACTCCGTCCTTTTGGAAACCAAGAGA 652
QY	663	TGCCATGAATTCCTTCAATACACTGTTGATGCTATGAGAAAGCATGCTTGAATGGCAG 722
DB		

Db	653	TGCCATGAATTCCTTCAATACACTGTTGATGCTATGACAGAAAGCATGCTTGAATGGCAG 712
QY	723	CAATAAATTAGACAGACACACCCAGGGCCACCACTCTTGTGTTGTGTCAGATATTTGGAGGATA 782
Db	713	CAATAAATTAGACAGACACACCCAGGGCCACCACTCTTGTGTTGTGTCAGATATTTGGAGGATA 772
QY	783	CCTAAGATCTAGAGTCAAAATGTTTAAATGCAAGGGCGTTTTCAGATATCTTTGATCCATA 842
Db	773	CCTAAGATCTAGAGTCAAAATGTTTAAATGCAAGGGCGTTTTCAGATATCTTTGATCCATA 832
QY	843	TCCTGATATAACATGAGAGATAAGGCTGCTCAGAGTGTCAAAGGCAATGGAGAGGATT 902
Db	833	TCCTGATATAACATGAGAGATAAGGCTGCTCAGAGTGTCAAAGGCAATGGAGAGGATT 892
QY	903	TGTGAAGCCGGAACAGCTTCATGGAGAGAAAACCTCGTCAAGTGCAGCAAGTGTAAAAAGAT 962
Db	893	TGTGAAGCCGGAACAGCTTCATGGAGAGAAAACCTCGTCAAGTGCAGCAAGTGTAAAAAGAT 952
QY	963	GGTTCAGCTTCAAGAGGTTCACTATCCATAGATCCTCTTAATGTTCTTACACTTTCTCT 1022
Db	953	GGTTCAGCTTCAAGAGGTTCACTATCCATAGATCCTCTTAATGTTCTTACACTTTCTCT 1012
QY	1023	GAACAGTTTTCGCAAAATTTTACCGGTGGAAGAAATGCTAAGGATGTGAAATACCTTGAGTA 1082
Db	1013	GAACAGTTTTCGCAAAATTTTACCGGTGGAAGAAATGCTAAGGATGTGAAATACCTTGAGTA 1072
QY	1083	TCTTGATATTCGGCCCATATATGTTCTCAACCCAGGAGAGCCAAATGTTCTACGTTCTTGTA 1142
Db	1073	TCTTGATATTCGGCCCATATATGTTCTCAACCCAGGAGAGCCAAATGTTCTACGTTCTTGTA 1132
QY	1143	TGCAGTGTCTGGTCCACACCTGGTTTTAAATGGCCATGTGGCCATTCTTGTCTACATAA 1202
Db	1133	TGCAGTGTCTGGTCCACACCTGGTTTTAAATGGCCATGTGGCCATTCTTGTCTACATAA 1192
QY	1203	AGCTAGCAATGCGCTCTGGTATCAATGAATGATGATCTCATTTGATCTACAGTGAATTTAG 1262
Db	1193	AGCTAGCAATGCGCTCTGGTATCAATGAATGATGATCTCATTTGATCTACAGTGAATTTAG 1252
QY	1263	ATCGGTACTCAGCAACAAGCCCTATGCTCTTTTATATATCAGTCCCATGATGTGAAAAA 1322
Db	1253	ATCGGTACTCAGCAACAAGCCCTATGCTCTTTTATATATCAGTCCCATGATGTGAAAAA 1312
QY	1323	TGGAGTGAACCTTACTCATCCACCCATAGCCCGGCGAGTCTCTCCCGCCCGCTCAT 1382
Db	1313	TGGAGTGAACCTTACTCATCCACCCATAGCCCGGCGAGTCTCTCCCGCCCGCTCAT 1372
QY	1383	CAGTCAGCGGGTGTCCACCAACAAACAGGCTCGGCCAGGCTTTATCGGACCAACAGTTCC 1442
Db	1373	CAGTCAGCGGGTGTCCACCAACAAACAGGCTCGGCCAGGCTTTATCGGACCAACAGTTCC 1432
QY	1443	CTCTCAGATGATAAAGAAATCCACCTCACTTAATGGGACTGGACCATTTGAAAGACACGCC 1502
Db	1433	CTCTCAGATGATAAAGAAATCCACCTCACTTAATGGGACTGGACCATTTGAAAGACACGCC 1492
QY	1503	AGCAGTTCATGTCGAGTCTTACGGGAAATTCAGTGTCAACAGGGCTAGTCTTGTAA 1562
Db	1493	AGCAGTTCATGTCGAGTCTTACGGGAAATTCAGTGTCAACAGGGCTAGTCTTGTAA 1552
QY	1563	TGCTTCAGCTTCTGTCCAAAACCTGGTCAGTTAAATAGGTCTCTCAGTATCCAGAACATCC 1622
Db	1553	TGCTTCAGCTTCTGTCCAAAACCTGGTCAGTTAAATAGGTCTCTCAGTATCCAGAACATCC 1612
QY	1623	TAAGAAAACAAAATAATCAATCAGTATTCACAAAGTTGCTGTTCCGCCAGTGTCACTC 1682
Db	1613	TAAGAAAACAAAATAATCAATCAGTATTCACAAAGTTGCTGTTCCGCCAGTGTCACTC 1672
QY	1683	TCAACCTTAACCTTCATAGTAATCTTTGGAGAAACCTTACCAAGCCCGTTCCTCTCTTAC 1742
Db	1673	TCAACCTTAACCTTCATAGTAATCTTTGGAGAAACCTTACCAAGCCCGTTCCTCTCTTAC 1732
QY	1743	CATTACCAATTCGCGGTACAGTCTACTCGAAACGATCTACGATGTCAAGTTCCTAGTAA 1802
Db	1733	CATTACCAATTCGCGGTACAGTCTACTCGAAACGATCTACGATGTCAAGTTCCTAGTAA 1792

Wed Aug 18 13:52:58 2004

QY	1803	ACTAACAAACCGATCCCGCCAGTGAATCTCTGCTCCCGAGCCGCTGATGAATGGCAAAATC	1862
Db	1793	AGTAACAAACCGATCCCGCCAGTGAATCTCTGCTCCCGAGCCGCTGATGAATGGCAAAATC	1852
QY	1863	CAAGCTGAATCTCAGCGTGTGTCCTTATGGCGCCGAGTCTCTCTGAGGACTCTCAACGA	1922
Db	1853	CAAGCTGAATCTCAGCGTGTGTCCTTATGGCGCCGAGTCTCTCTGAGGACTCTCAACGA	1912
QY	1923	GGAGTCAAAAGGGCTCGGCAAGGAGAGATGGATTTGGTACGATTTGAGCTCCCACTCTCC	1982
Db	1913	GGAGTCAAAAGGGCTCGGCAAGGAGAGATGGATTTGGTACGATTTGAGCTCCCACTCTCC	1972
QY	1983	CGGCCAAGATGCCGAAGATGAGGAGGCCACTCCGACAGAGCTTCAAGAACCCATGACCCCT	2042
Db	1973	CGGCCAAGATGCCGAAGATGAGGAGGCCACTCCGACAGAGCTTCAAGAACCCATGACCCCT	2032
QY	2043	AAACGGTGTCTAATAGTGCAGACAGCAGTGAACCTGACCCGAAAGAAAGCGCTAGCGCTGA	2102
Db	2033	AAACGGTGTCTAATAGTGCAGACAGCAGTGAACCTGACCCGAAAGAAAGCGCTAGCGCTGA	2092
QY	2103	TGCTGCCAGCTGCCAAGGCCAGCTGCTCCCTGCACTCAGAAAAATCCCTTTTCTAAGGCAAA	2162
Db	2093	TGCTGCCAGCTGCCAAGGCCAGCTGCTCCCTGCACTCAGAAAAATCCCTTTTCTAAGGCAAA	2152
QY	2163	CGGTCTTCTCGAAAGTTGATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2222
Db	2153	CGGTCTTCTCGAAAGTTGATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2212
QY	2223	AGAGACCTTCAGCTTAGCAACAACTGAAGGCTCGACGAGTGAATGAGTGCACCTGG	2282
Db	2213	AGAGACCTTCAGCTTAGCAACAACTGAAGGCTCGACGAGTGAATGAGTGCACCTGG	2272
QY	2283	AGCAGAGAGGGCCCTCCGAGGACCGGACCGAGCCTCAGCCTCGCAGCCCGCCCGC	2342
Db	2273	AGCAGAGAGGGCCCTCCGAGGACCGGACCGAGCCTCAGCCTCGCAGCCCGCCCGC	2332
QY	2343	CGAATCCCTCGAGGAGCGAGTGGCGGC	2372
Db	2333	CGAATCCCTCGAGGAGCGAGTGGCGGC	2362
RESULT 8			
AX364931			
LOCUS			
Sequence 82 from Patent WO0206315.			
AX364931			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Homo sapiens (human)			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE			
Mintz, L., Freilich, S. and Bernstein, J.			
Novel nucleic acid and amino acid sequences			
Patent: WO 0206315-A 82 24-JAN-2002;			
CompuGen Ltd. (IL)			
FEATURES			
Location/Qualifiers			
1..1679			
/organism="Homo sapiens"			
/mol_type="unassigned DNA"			
/db_xref="taxon:9606"			
ORIGIN			
Query Match			
Best Local Similarity 61.2%; Score 1475; DB 6; Length 1679;			
Matches 1675; Conservative 0; Mismatches 4; Indels 0; Gaps 0;			
QY			
694 GCTATGCGAAAGCATGCTTGAATGGCAGCAATAAATAGACAGACACCCAGGCCACC			
753			
Db			
1 GCTATGCGAAAGCATGCTTGAATGGCAGCAATAAATAGACAGACACCCAGGCCACC			
60			

QY	754	ACTCTTGTGTCAGATATTTGGAGATACCTAAGATCTAGAGTCAAAATGTTTAAATTC	813
Db	61	ACTCTTGTGTCAGATATTTGGAGATACCTAAGATCTAGAGTCAAAATGTTTAAATTC	120
QY	814	AAGGGCGTTTCAGATATTTGATCCATATCTTGATATTAACATTTGAGATAAAGGCTGCT	873
Db	121	AAGGGCGTTTCAGATATTTGATCCATATCTTGATATTAACATTTGAGATAAAGGCTGCT	180
QY	874	CAGAGTGTCAACAGGCAATTTGGAGCAGTTTGTGAAGCCGGAACAGCTTGATGAGAGAAAC	933
Db	181	CAGAGTGTCAACAGGCAATTTGGAGCAGTTTGTGAAGCCGGAACAGCTTGATGAGAGAAAC	240
QY	934	TGCTACAAGTGCAGCAAGTGTAAAAAGATGGTTCCAGCTTCAAAAGAGTTCACTATCCAT	993
Db	241	TGCTACAAGTGCAGCAAGTGTAAAAAGATGGTTCCAGCTTCAAAAGAGTTCACTATCCAT	300
QY	994	AGATCCCTCTAATGTTCTTACACTTTCTGTAACGTTTTCGCAAAATTTACCGGTGGAATA	1053
Db	301	AGATCCCTCTAATGTTCTTACACTTTCTGTAACGTTTTCGCAAAATTTACCGGTGGAATA	360
QY	1054	ATTGCTAAGGATGTAATAACCTGAGTATCTTTGATATTCGGCCATATATGTCTCAACCC	1113
Db	361	ATTGCTAAGGATGTAATAACCTGAGTATCTTTGATATTCGGCCATATATGTCTCAACCC	420
QY	1114	AACGGAGAGCAATTTGTCTACGTTCTGATGCAAGTGTGTCACACTGGTTTAAATTC	1173
Db	421	AACGGAGAGCAATTTGTCTACGTTCTGATGCAAGTGTGTCACACTGGTTTAAATTC	480
QY	1174	CATGCTGGCCATTTACTTCTGCTACATAAAGCTTAGCAATGSCCTCTGATCAATGAAT	1233
Db	481	CATGCTGGCCATTTACTTCTGCTACATAAAGCTTAGCAATGSCCTCTGATCAATGAAT	540
QY	1234	GACTCCATTTGATCTACAGTGTATTTAGATCGGTACTCAGCCCAACAGCCTATGTGCTC	1293
Db	541	GACTCCATTTGATCTACAGTGTATTTAGATCGGTACTCAGCCCAACAGCCTATGTGCTC	600
QY	1294	TTTTATATCAGGTCCTCCTGATGTAAGAAATGGAGGTGAATTAATCTATCCACCCTAGC	1353
Db	601	TTTTATATCAGGTCCTCCTGATGTAAGAAATGGAGGTGAATTAATCTATCCACCCTAGC	660
QY	1354	CCCGGCCAGTCTCTCCCGCCCGCTCATCAGTCAGGGGTTGTCACCAACAAACAGGCT	1413
Db	661	CCCGGCCAGTCTCTCTCCCGCCCGCTCATCAGTCAGGGGTTGTCACCAACAAACAGGCT	720
QY	1414	CGCGCCAGCTTTATCGGACCAACAGCTTCCCTCTCACATGATAAAGAAATCCACCTCATT	1473
Db	721	CGCGCCAGCTTTATCGGACCAACAGCTTCCCTCTCACATGATAAAGAAATCCACCTCATT	780
QY	1474	AATGGGACTGGACCAATTTGAAAGACACGCAAGCAGTTCCATGTGAGTCCCTAAACGGG	1533
Db	781	AATGGGACTGGACCAATTTGAAAGACACGCAAGCAGTTCCATGTGAGTCCCTAAACGGG	840
QY	1534	TCCAGTGTCAACAGGCTAGTCTCTGTTAATGCTTCAGCTTCTGTCACCAACCTGTCAGT	1593
Db	841	TCCAGTGTCAACAGGCTAGTCTCTGTTAATGCTTCAGCTTCTGTCACCAACCTGTCAGT	900
QY	1594	AATAGTCTCAGTGTATCCAGAACATCCTTAAGAAACAAAAAATTAACATCAGTATTCAC	1653
Db	901	AATAGTCTCAGTGTATCCAGAACATCCTTAAGAAACAAAAAATTAACATCAGTATTCAC	960
QY	1654	AACAGTGTCTGTTGCGCCAGTGTCACTCAACCTAACCCTTCAGTAAATTTCTTGGAG	1713
Db	961	AACAGTGTCTGTTGCGCCAGTGTCACTCAACCTAACCCTTCAGTAAATTTCTTGGAG	1020
QY	1714	AACCTTACCAAGCCGTTTCCCTCTTCTACCAATTTACCAATTTCTGAGTACAGTCTACCTC	1773
Db	1021	AACCTTACCAAGCCGTTTCCCTCTTCTACCAATTTACCAATTTCTGAGTACAGTCTACCTC	1080
QY	1774	AACGATCTACGATCTCAGTCTTCTAGTAAAGTAAACAAACCGATCCCCCGAGTGAATCC	1833
Db	1081	AACGATCTACGATCTCAGTCTTCTAGTAAAGTAAACAAACCGATCCCCCGAGTGAATCC	1140
QY	1834	TGCTCCAGCCCGCTGATGAATGGCAAAATCCAAAGTGAATCCAGCGTGTGGTCCCTAT	1893

```
Db      1141  |||||TGCTCCAGCCCGTGATGAATGGCAATCCAAAGCTGAATCCAGCGTCTGGTCCCTAT 1200
QY      1894  GGGCCCGAGTCTCTGAGGAGTCTGAGCAGAGTCAAGGGGCTGGCAGAGGAATGGG 1953
Db      1201  GGGCCCGAGTCTCTGAGGAGTCTGAGCAGAGTCAAGGGGCTGGCAGAGGAATGGG 1260
QY      1954  ATTGGTACGATTGTGAGCTCCCACTCTCCCGGCCCAAGATGCCGAAGATGAGGAGGCCACT 2013
Db      1261  ATTGGTACGATTGTGAGCTCCCACTCTCCCGGCCCAAGATGCCGAAGATGAGGAGGCCACT 1320
QY      2014  CCGCAGAGCTTCAAGAACCATGACCTTAACCGGTCTAATAGTGACAGACGACAGT 2073
Db      1321  CCGCAGAGCTTCAAGAACCATGACCTTAACCGGTCTAATAGTGACAGACGACAGT 1380
QY      2074  GACCCGAGAAAGAAACGGCTAGCGCTGATGGTCCAGCTGCCAAGGCCAGCTGCCCTG 2133
Db      1381  GACCCGAGAAAGAAACGGCTAGCGCTGATGGTCCAGCTGCCAAGGCCAGCTGCCCTG 1440
QY      2134  CACTCAGAAATCCCTTTGCTAAGGCAAAACGGTCTTCTCGAAAGTTGATGCTGCTCCT 2193
Db      1441  CACTCAGAAATCCCTTTGCTAAGGCAAAACGGTCTTCTCGAAAGTTGATGCTGCTCCT 1500
QY      2194  TTGCTCTCTCTCCAGAGCAAAATCTTAGACACCTTCAGGCTTAGCAACAACTGAA 2253
Db      1501  TTGCTCTCTCTCCAGAGCAAAATCTTAGACACCTTCAGGCTTAGCAACAACTGAA 1560
QY      2254  GGCTCAGCGATGAATAGTGCACCTGAGCAGAGAGGGCCCTCCGAGGACCGCGAC 2313
Db      1561  GGCTCAGCGATGAATAGTGCACCTGAGCAGAGAGGGCCCTCCGAGGACCGCGAC 1620
QY      2314  GCGAGCCTCAGCTGGCGAGCCCGCCGCGGAACTCCCTGGAGGAGCAGATGCGCGCG 2372
Db      1621  GCGAGCCTCAGCTGGCGAGCCCGCCGCGGAACTCCCTGGAGGAGCAGATGCGCGCG 1679
```

```
RESULT 9
AC004895/c
LOCUS      152927 bp      DNA      linear      PRI 30-SEP-2000
DEFINITION Homo sapiens PAC clone RP4-810E6 from 7, complete sequence.
ACCESSION AC004895
VERSION    AC004895.2 GI:4926908
KEYWORDS   HTG.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 152927)
AUTHORS    Sulston,J.E. and Waterston,R.
TITLE      Toward a complete human genome sequence
JOURNAL    Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE    99063792
PUBMED     9847074
REFERENCE  2 (bases 1 to 152927)
AUTHORS    Cordes,M. and Gibson,A.
TITLE      The sequence of Homo sapiens PAC clone RP4-810E6
JOURNAL    Unpublished
REFERENCE  3 (bases 1 to 152927)
AUTHORS    Waterston,R.H.
TITLE      Direct Submission
JOURNAL    Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE  4 (bases 1 to 152927)
AUTHORS    Waterston,R.H.
TITLE      Direct Submission
JOURNAL    Submitted (29-MAY-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE  5 (bases 1 to 152927)
AUTHORS    Waterston,R.
TITLE      Direct Submission
JOURNAL    Submitted (30-SEP-2000) Department of Genetics, Washington
```

COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 29, 1999 this sequence version replaced gi:3450909.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_DJ0810E06

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GTB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.

VECTOR: PCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP1-42M2. Actual start of this clone is at base position 1 of RP4-810E6; actual end is at 152927 of RP4-810E6.

There is a questionable number of A's from 81574-81601. The consensus reflects 27 A's but could be as many as 29.

FEATURES
source

```
1. 152927
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP4-810E6"
/clone_lib="RPCI-4"
329..464
/rpt_family="L1"
466..560
/rpt_family="MIR"
649..833
/rpt_family="Alu"
1269..1585
/rpt_family="Alu"
1586..1745
/rpt_family="L1"
1813..1974
/rpt_family="Alu"
2887..2970
```

```
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
```

repeat_region	16336..6460	/rpt_family="Alu"	16336..6460	repeat_region	16336..6460	/rpt_family="Alu"	16336..6460
repeat_region	16480..6693	/rpt_family="L1"	16480..6693	repeat_region	16480..6693	/rpt_family="L1"	16480..6693
repeat_region	16793..17187	/rpt_family="Alu"	16793..17187	repeat_region	16793..17187	/rpt_family="Alu"	16793..17187
repeat_region	17675..17877	/rpt_family="L1"	17675..17877	repeat_region	17675..17877	/rpt_family="L1"	17675..17877
misc_feature	18340..18452	/note="match to EST M79246 (NID:g273559)"	18340..18452	misc_feature	18340..18452	/note="match to EST M79246 (NID:g273559)"	18340..18452
misc_feature	18400..18458	/note="similar to EST AA284549 (NID:g1927478) zt24f01.xl"	18400..18458	misc_feature	18400..18458	/note="similar to EST AA284549 (NID:g1927478) zt24f01.xl"	18400..18458
repeat_region	19073..19720	/rpt_family="L1"	19073..19720	repeat_region	19073..19720	/rpt_family="L1"	19073..19720
repeat_region	19721..20026	/rpt_family="Alu"	19721..20026	repeat_region	19721..20026	/rpt_family="Alu"	19721..20026
repeat_region	20027..20435	/rpt_family="L1"	20027..20435	repeat_region	20027..20435	/rpt_family="L1"	20027..20435
repeat_region	20436..20735		20436..20735	repeat_region	20436..20735		20436..20735
Query Match 29.9%; Score 722; DB 9; Length 152927; Best Local Similarity 100.0%; Pred. No. 0; Matches 722; Conservative 0; Mismatches 0; Indels 0; Gaps 0;							
Qy	1457	AGAATCCACCTCACTTAATGGGACTGGACCATTTGAAGACAGCCAGCAGTTCCATGT	1516	Qy	1457	AGAATCCACCTCACTTAATGGGACTGGACCATTTGAAGACAGCCAGCAGTTCCATGT	1516
Db	46699	AGAATCCACCTCACTTAATGGGACTGGACCATTTGAAGACAGCCAGCAGTTCCATGT	46640	Db	46699	AGAATCCACCTCACTTAATGGGACTGGACCATTTGAAGACAGCCAGCAGTTCCATGT	46640
Qy	1517	CGAGTCTTAACGGGATTCAGTGTCAACAGGCGTAGTCTGTAAATGCTTACGTTCTG	1576	Qy	1517	CGAGTCTTAACGGGATTCAGTGTCAACAGGCGTAGTCTGTAAATGCTTACGTTCTG	1576
Db	46639	CGAGTCTTAACGGGATTCAGTGTCAACAGGCGTAGTCTGTAAATGCTTACGTTCTG	46580	Db	46639	CGAGTCTTAACGGGATTCAGTGTCAACAGGCGTAGTCTGTAAATGCTTACGTTCTG	46580
Qy	1577	TCCAAATCTGTCTAGTTAATAGTCTCTCAGTGATCCAGAACATCTTAAGAAACAAAA	1636	Qy	1577	TCCAAATCTGTCTAGTTAATAGTCTCTCAGTGATCCAGAACATCTTAAGAAACAAAA	1636
Db	46579	TCCAAATCTGTCTAGTTAATAGTCTCTCAGTGATCCAGAACATCTTAAGAAACAAAA	46520	Db	46579	TCCAAATCTGTCTAGTTAATAGTCTCTCAGTGATCCAGAACATCTTAAGAAACAAAA	46520
Qy	1637	TTCAATCAGTATTCACAAAGTTGCTGTTCGCCAGTGTCTCAGTCTCAACCTAACCTTC	1696	Qy	1637	TTCAATCAGTATTCACAAAGTTGCTGTTCGCCAGTGTCTCAGTCTCAACCTAACCTTC	1696
Db	46519	TTCAATCAGTATTCACAAAGTTGCTGTTCGCCAGTGTCTCAGTCTCAACCTAACCTTC	46460	Db	46519	TTCAATCAGTATTCACAAAGTTGCTGTTCGCCAGTGTCTCAGTCTCAACCTAACCTTC	46460
Qy	1697	ATAGTAATCTTTGGAGAACCTTACCAAGCCGTTCCCTCTCTACCATTTACCAATCTG	1756	Qy	1697	ATAGTAATCTTTGGAGAACCTTACCAAGCCGTTCCCTCTCTACCATTTACCAATCTG	1756
Db	46459	ATAGTAATCTTTGGAGAACCTTACCAAGCCGTTCCCTCTCTACCATTTACCAATCTG	46400	Db	46459	ATAGTAATCTTTGGAGAACCTTACCAAGCCGTTCCCTCTCTACCATTTACCAATCTG	46400
Qy	1757	CAGTACAGTCTACCTCGAAAGCATCTACGATGTCAAGTTCTAGTAAAGTAAACAAACCGA	1816	Qy	1757	CAGTACAGTCTACCTCGAAAGCATCTACGATGTCAAGTTCTAGTAAAGTAAACAAACCGA	1816
Db	46399	CAGTACAGTCTACCTCGAAAGCATCTACGATGTCAAGTTCTAGTAAAGTAAACAAACCGA	46340	Db	46399	CAGTACAGTCTACCTCGAAAGCATCTACGATGTCAAGTTCTAGTAAAGTAAACAAACCGA	46340
Qy	1817	TCCCGCCGAGTGAATCTGTCTCCAGCCCGTGAATGGCAAAATCCAAAGCTGAATCCA	1876	Qy	1817	TCCCGCCGAGTGAATCTGTCTCCAGCCCGTGAATGGCAAAATCCAAAGCTGAATCCA	1876
Db	46339	TCCCGCCGAGTGAATCTGTCTCCAGCCCGTGAATGGCAAAATCCAAAGCTGAATCCA	46280	Db	46339	TCCCGCCGAGTGAATCTGTCTCCAGCCCGTGAATGGCAAAATCCAAAGCTGAATCCA	46280
Qy	1877	CGGTGCTGTGCTGCTTATGGCCCGAGTCTCTGAGGACTCTGACGAGGAGTCAAAAGGGGC	1936	Qy	1877	CGGTGCTGTGCTGCTTATGGCCCGAGTCTCTGAGGACTCTGACGAGGAGTCAAAAGGGGC	1936
Db	46279	CGGTGCTGTGCTGCTTATGGCCCGAGTCTCTGAGGACTCTGACGAGGAGTCAAAAGGGGC	46220	Db	46279	CGGTGCTGTGCTGCTTATGGCCCGAGTCTCTGAGGACTCTGACGAGGAGTCAAAAGGGGC	46220
Qy	1937	TGGCAAGGAGAAATGGGATTTGGTACGATTTGAGCTTCCCACTCTCCCGCCCAAGATGCG	1996	Qy	1937	TGGCAAGGAGAAATGGGATTTGGTACGATTTGAGCTTCCCACTCTCCCGCCCAAGATGCG	1996
Db	46219	TGGCAAGGAGAAATGGGATTTGGTACGATTTGAGCTTCCCACTCTCCCGCCCAAGATGCG	46160	Db	46219	TGGCAAGGAGAAATGGGATTTGGTACGATTTGAGCTTCCCACTCTCCCGCCCAAGATGCG	46160
Qy	1937	AAGATGAGGAGCCACTCCGACGAGCTTCAAGAACCATGACCTTAACGGTCTGAATA	2056	Qy	1937	AAGATGAGGAGCCACTCCGACGAGCTTCAAGAACCATGACCTTAACGGTCTGAATA	2056
Db	46159	AAGATGAGGAGCCACTCCGACGAGCTTCAAGAACCATGACCTTAACGGTCTGAATA	46100	Db	46159	AAGATGAGGAGCCACTCCGACGAGCTTCAAGAACCATGACCTTAACGGTCTGAATA	46100
Qy	2057	GTGCAGACGAGCAGTGTACCGAAAGAAACCGCTAGCGCTGATGTGCCAGCTGCC	2116	Qy	2057	GTGCAGACGAGCAGTGTACCGAAAGAAACCGCTAGCGCTGATGTGCCAGCTGCC	2116
Db	46099	GTGCAGACGAGCAGTGTACCGAAAGAAACCGCTAGCGCTGATGTGCCAGCTGCC	46040	Db	46099	GTGCAGACGAGCAGTGTACCGAAAGAAACCGCTAGCGCTGATGTGCCAGCTGCC	46040
Qy	2117	AAGGCCAGCTGCTTGTGCACTCAGAAATCCCTTTGCTTAAGGCAACCGTCTTCTGGAA	2176	Qy	2117	AAGGCCAGCTGCTTGTGCACTCAGAAATCCCTTTGCTTAAGGCAACCGTCTTCTGGAA	2176
Db	46039	AAGGCCAGCTGCTTGTGCACTCAGAAATCCCTTTGCTTAAGGCAACCGTCTTCTGGAA	45980	Db	46039	AAGGCCAGCTGCTTGTGCACTCAGAAATCCCTTTGCTTAAGGCAACCGTCTTCTGGAA	45980
Qy	2177	AG 2178		Qy	2177	AG 2178	

```

Db      45979 AG 45978
||
RESULT 10
AC069393
LOCUS   154791 bp DNA linear HTG 26-OCT-2000
DEFINITION Homo sapiens chromosome 7 clone RP11-585K18, WORKING DRAFT
SEQUENCE, 14 unordered pieces.
AC069393
ACCESSION AC069393.8 GI:11024947
VERSION   HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS  Homo sapiens (human)
SOURCE    Homo sapiens
ORGANISM  Homo sapiens
          Bkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          Waterston,R.H.
          The sequence of Homo sapiens clone
REFERENCE 2 (bases 1 to 154791)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (28-MAY-2000) Genome Sequencing Center, Washington
          University School of Medicine, 444 Forest Park Parkway, St. Louis,
          MO 63108, USA
COMMENT   On Oct 26, 2000 this sequence version replaced gi:10946568.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0585K18
----- Summary Statistics -----
Sequencing vector: M13; 71%
Sequencing vector: plasmid; 5%
Chemistry: Dye-primer ET; 71% of reads
Assembly: Dye-terminator Big Dye; 5% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 143345 bases at least Q40
Consensus quality: 146572 bases at least Q30
Consensus quality: 148386 bases at least Q20
Insert size: 164000; agarose-fp
Insert size: 153491; sum-of-contigs
Quality coverage: 4.65 in Q20 bases; agarose-fp
Quality coverage: 5.04 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      5209: contig of 5209 bp in length
*      5210: gap of unknown length
*      5310: contig of 4762 bp in length
*      10071: gap of unknown length
*      10171: gap of unknown length
*      10172: contig of 5993 bp in length
*      16165: gap of unknown length
*      16265: gap of unknown length
*      29394: contig of 13130 bp in length
*      29395: gap of unknown length
*      29495: contig of 12096 bp in length
*      41591: gap of unknown length
*      41691: gap of unknown length
*      54703: contig of 13013 bp in length
*      54704: gap of unknown length
*      74381: contig of 19578 bp in length
*      74481: gap of unknown length
*      103253: contig of 28772 bp in length
*      103254: gap of unknown length
*      139431: contig of 36078 bp in length

```

```

*      139432 139531: gap of unknown length
*      139532 140718: contig of 1187 bp in length
*      140719 140818: gap of unknown length
*      140819 143471: contig of 2653 bp in length
*      143472 143571: gap of unknown length
*      143572 147062: contig of 3491 bp in length
*      147063 147162: gap of unknown length
*      147163 150060: contig of 2898 bp in length
*      150061 150160: gap of unknown length
*      150161 154791: contig of 4631 bp in length.
FEATURES
        Location/Qualifiers
            1..154791
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="7"
            clone="RP11-585K18"
            1..5209
                /note="assembly_name:Contig10"
            5310..10071
                /note="assembly_name:Contig11"
            10172..16164
                /note="assembly_name:Contig12"
            clone_end:77
            vector_side:left
            16265..29394
                /note="assembly_name:Contig13"
            29495..41590
                /note="assembly_name:Contig14"
            41691..54703
                /note="assembly_name:Contig15"
            54804..74381
                /note="assembly_name:Contig16"
            74482..103253
                /note="assembly_name:Contig17"
            103354..139431
                /note="assembly_name:Contig18"
            139532..140718
                /note="assembly_name:Contig19"
            clone_end:SP6
            vector_side:right
            140819..143471
                /note="assembly_name:Contig6"
            143572..147062
                /note="assembly_name:Contig7"
            147163..150060
                /note="assembly_name:Contig8"
            150161..154791
                /note="assembly_name:Contig9"
ORIGIN
Query Match      29.9%; Score 722; DB 2; Length 154791;
Best Local Similarity 100.0%; Pred.No; 0;
Matches 722; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1457 AGAATCCACTCCTTAATAATGGGACTGGACCATTAAGAGACACGCCAAGAGATTCCCATGT 1516
Db      129201 AGAATCCACTCCTTAATAATGGGACTGGACCATTAAGAGACACGCCAAGAGATTCCCATGT 129260
QY      1517 CGAGTCTCTACAGGGAATTCAGTGTCAACAGGGCTAGTCTCTGTTAATGCTCAGCTTCTG 1576
Db      129261 CGAGTCTCTACAGGGAATTCAGTGTCAACAGGGCTAGTCTCTGTTAATGCTCAGCTTCTG 129320
QY      1577 TCCAAAACCTGGTCAGTTAATAGTCTCTCAGTGATCCAGACATCCCTAAGAAACAAAAA 1636
Db      129321 TCCAAAACCTGGTCAGTTAATAGTCTCTCAGTGATCCAGACATCCCTAAGAAACAAAAA 129380
QY      1637 TTCAATCAGTATTCAACAAGTTGCTGTTCCGCGAGTGTGAGTCTCAACCTTAACCTTC 1696
Db      129381 TTCAATCAGTATTCAACAAGTTGCTGTTCCGCGAGTGTGAGTCTCAACCTTAACCTTC 129440
QY      1697 ATAGTAATCTTTGGAGAACCCCTACCAAGCCGTTCCCTCTTACCAATTACCAATTCG 1756

```



```
/note="assembly_name:Contig25"
5281..7152
/note="assembly_name:Contig26"
7253..9449
/note="assembly_name:Contig27"
9550..10943
/note="assembly_name:Contig28"
11044..113812
/note="assembly_name:Contig29"
13913..18353
/note="assembly_name:Contig30"
18454..21675
/note="assembly_name:Contig31"
21776..24307
/note="assembly_name:Contig32"
24408..28212
/note="assembly_name:Contig33"
28113..34694
/note="assembly_name:Contig34"
34795..38278
/note="assembly_name:Contig35"
38379..44872
/note="assembly_name:Contig36"
44973..51522
/note="assembly_name:Contig37"
51623..59330
/note="assembly_name:Contig38"
59431..67685
/note="assembly_name:Contig39"
67786..76304
/note="assembly_name:Contig40"
76405..91684
/note="assembly_name:Contig41"
91785..103781
/note="assembly_name:Contig42"
103882..115983
/note="assembly_name:Contig43"
116084..131828
/note="assembly_name:Contig44"
131929..148774
/note="assembly_name:Contig45"
148875..168325
/note="assembly_name:Contig46"
168426..185595
/note="assembly_name:Contig47"
185696..217022
/note="assembly_name:Contig48"

ORIGIN
Query Match 17.2%; Score 415; DB 2; Length 217022;
Best Local Similarity 99.5%; Pred. No. 2.1e-230;
Matches 565; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1457 AGAATCCACCTCACTTAATCGGACTGACCATTAAGAAAGACACGCAAGCAGTTCATGT 1516
DB 150471 AGAATCCACCTCACTTAATCGGACTGACCATTAAGAAAGACACGCAAGCAGTTCATGT 150412

QY 1517 CGAGTCCCTAACGGGAATCCACAGTGTCAACAGGGCTAGTCCCTGTTAATGCTTCAGCTTCG 1576
DB 150411 CGAGTCCCTAACGGGAATTCAGTGTCAACAGGGCTAGTCCCTGTTAATGCTTCAGCTTCG 150352

QY 1577 TCCAAAACCTGTCAGTTAATAGGTCCTCAGTGTATCCAGAACATCCTTAAGAAACAAAAA 1636
DB 150351 TCCAAAACCTGTCAGTTAATAGGTCCTCAGTGTATCCAGAACATCCTTAAGAAACAAAAA 150292

QY 1637 TTCAATCAGTATTCACAAAGTTGCTGTTCGCCAGTGTGTCAGTCTCAACCTAACCTTC 1696
DB 150291 TTCAATCAGTATTCACAAAGTTGCTGTTCGCCAGTGTGTCAGTCTCAACCTAACCTTC 150232

QY 1697 ATAGTAATCTTTGGAGAACCTTACCAAGCCCGTTCCCTCTTCTACCAATTCATTCG 1756
DB 150231 ATAGTAATCTTTGGAGAACCTTACCAAGCCCGTTCCCTCTTCTACCAATTCATTCG 150172
```

```
QY 1757 CAGTACAGTCTACCTCGAACGCATCTACGATGTCTAGTAAAGTAACTAAACAAACCGA 1816
DB 150171 CAGTACAGTCTACCTCGAACGCATCTACGATGTCTAGTAAAGTAACTAAACAAACCGA 150112

QY 1817 TCCCCCGCAGTGAAATCCTGCTCCAGCCGCGTGATGAAATGCAAAATCCAAGCTGAACCTCA 1876
DB 150111 TCCCCCGCAGTGAAATCCTGCTCCAGCCGCGTGATGAAATGCAAAATCCAAGCTGAACCTCA 150052

QY 1877 GCCTGTCTGTGCCCTATGCGCGCCGAGTCTCTCTGAGGACTCTGACGAGAGTCAAGGGGC 1936
DB 150051 GCCTGTCTGTGCCCTATGCGCGCCGAGTCTCTCTGAGGACTCTGACGAGAGTCAAGGGGC 149992

QY 1937 TGGCGAAGGAGAAATGGGAATGGTACGATTGTGAGCTCCCACTCTCCCGGCCCAAGATCCCG 1996
DB 149991 TGGCGAAGGAGAAATGGGAATGGTACGATTGTGAGCTCCCACTCTCCCGGCCCAAGATCCCG 149932

QY 1997 AAGATGAGGAGGCCACTCCGCACGAGCT 2024
DB 149931 AAGATGAGGAGGCCACTCCGCACGAGCT 149904

RESULT 12
AR264045
LOCUS AR264045 722 bp DNA linear PAT 29-JAN-2003
DEFINITION Sequence 223 from patent US 6331427.
ACCESSION AR264045
VERSION AR264045.1 GI:28076049
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 722)
TITLES Robison, K.E.
JOURNAL Protease homologs
FEATURES Patent: US 6331427-A 223 18-DEC-2001;
Location/Qualifiers
source 1..722
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 14.8%; Score 357; DB 6; Length 722;
Best Local Similarity 100.0%; Pred. No. 1.5e-196;
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1061 AGGATGTCAAATACCTCAGTATCTTGATATCGGCCATATATGTCTCAACCAACGGAG 1120
DB 107 AGGATGTCAAATACCTCAGTATCTTGATATCGGCCATATATGTCTCAACCAACGGAG 166

QY 1121 AGCCAATGTCTACGTCTTGATGTCAGTGTGTCACACTGGTTTAAATGCAATGCTG 1180
DB 167 AGCCAATGTCTACGTCTTGATGTCAGTGTGTCACACTGGTTTAAATGCAATGCTG 226

QY 1181 GCCATTACTTCTGCTACATAAAGCTAGCAATGCGCTCTGGTATCAATGAATGACTCCA 1240
DB 227 GCCATTACTTCTGCTACATAAAGCTAGCAATGCGCTCTGGTATCAATGAATGACTCCA 286

QY 1241 TTGTATCTACCAAGTATATAGATCGGTACTCAGCAACAAGCCTATGTGCTCTTTATA 1300
DB 287 TTGTATCTACCAAGTATATAGATCGGTACTCAGCAACAAGCCTATGTGCTCTTTATA 346

QY 1301 TCAGGTCCTCATGTGAAAAAATGGAGTGAACCTTACTCATCCACCCATAGCCCGGCC 1360
DB 347 TCAGGTCCTCATGTGAAAAAATGGAGTGAACCTTACTCATCCACCCATAGCCCGGCC 406

QY 1361 AGTCCTCTCCCGCCCGCTCATCAGTCAGCGGGTGTGTCACCAACAAACAGCTGCGC 1417
DB 407 AGTCCTCTCCCGCCCGCTCATCAGTCAGCGGGTGTGTCACCAACAAACAGCTGCGC 463

RESULT 13
AC146146
LOCUS AC146146 217022 bp DNA linear HTG 01-AUG-2003
```


[illegible]

<hr/>					
ACCESSION	BD147651				
VERSION	BD147651.1	GI:27853409			
KEYWORDS	JP 2002191363-A/2494				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
<hr/>					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
<hr/>					
TITLE	Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,				
JOURNAL	Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.				
<hr/>					
COMMENT	Primer for synthesizing full-length cDNA and use thereof				
	Patent: JP 2002191363-A 2494 09-JUL-2002;				
	HELIX RESEARCH INSTITUTE				
	OS Homo sapiens (human)				
	PN JP 2002191363-A/2494				
	PD 09-JUL-2002				
	PF 28-JUL-2000 JP 2000280990				
	PI TOGHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU				
	PI SAITO,				
	PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,				
	PI KENICHI NAGAI, TETSUOI OTSUKI				
	PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC				
	10,C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC				
	PC C12P21/02,C12O1/68/C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC				
	Primer for synthesizing full-length cDNA and use thereof FH Key				
	FT source 1..521				
	FT Location/Qualifiers /organism='Homo sapiens (human)'. '				
FEATURES	source 1..521				
	Location/Qualifiers /organism='Homo sapiens'				
	/mol_type='genomic DNA'				
	/db_xref='taxon:9606'				
ORIGIN					
	Query Match 9.6%; Score 232; DB 6; Length 521;				
	Best Local Similarity 99.6%; Pred. No. 2.7e-123;				
	Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
QY	6 GTGTGCGCGGGCGGGCGGGCGGGCGGGATGGAGCGGCGGAGCCGGCGGTCAGAG 65				
Dd	1 GTGTGCGCGGGCGGGCGGGCGGGCGGGATGGAGCGGCGGAGCCGGCTCAGAG 60				
QY	66 TTGAACAATGACCATTAGTTGACAAAGCTTCTGAATCTTCAGACCATCAGCCTATCAAA 125				
Dd	61 TTGAACAATGACCATTAGTTGACAAAGCTTCTGAATCTTCAGACCATCAGCCTATCANAA 120				
QY	126 TCAGCCTGCAGCTCCGAGGCGAGTCTCACCTGGAGACATGGATGAGGTTCGCCAGCTG 185				
Dd	121 TCAGCCTGCAGCTCCGAGGCGAGTCTCACCTGGAGACATGGATGAGGTTCGCCAGCTG 180				
QY	186 GGGTGCTGTGCTTTCATTGAATGATGTGCAAAATCACACACTTTCTTTAGGACCGAGTACC 245				
Dd	181 GGGTGCTGTGCTTTCATTGAATGATGTGCAAAATCACACACTTTCTTTAGGACCGAGTACC 240				
QY	246 TGGTGCTGTAGTTTATTTCAGTTTCATCTCTGCTACCTGATAAATCA 288				
Dd	241 TGGTGCTGTAGTTTATTTCAGTTTCATCTCTGCTACCTGATAAATCA 283				
<hr/>					
RESULT 17					
LOCUS	AR264036	457 bp	DNA	linear	PAT 29-JAN-2000
DEFINITION	Sequence 214 from patent US 6331427.				
ACCESSION	AR264036				
VERSION	AR264036.1	GI:28076040			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 457)				
AUTHORS	Robison,K.E				

* NOTE: This is a 'working draft' sequence. It currently consists of 14 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

* 1 5209: contig of 5209 bp in length
* 5210: gap of unknown length
* 5310: contig of 4762 bp in length
* 10072: contig of 4762 bp in length
* 10171: gap of unknown length
* 10172: contig of 5993 bp in length
* 16164: gap of unknown length
* 16165: contig of 13130 bp in length
* 16265: gap of unknown length
* 29395: gap of unknown length
* 29495: contig of 12096 bp in length
* 41591: gap of unknown length
* 41592: gap of unknown length
* 41690: gap of unknown length
* 54703: contig of 13013 bp in length
* 54704: gap of unknown length
* 54804: contig of 19578 bp in length
* 74382: gap of unknown length
* 74482: contig of 28772 bp in length
* 103254: gap of unknown length
* 103255: gap of unknown length
* 139432: contig of 36078 bp in length
* 139433: gap of unknown length
* 139532: contig of 1187 bp in length
* 140719: gap of unknown length
* 140818: gap of unknown length
* 143472: contig of 2653 bp in length
* 143473: gap of unknown length
* 147062: contig of 3491 bp in length
* 147063: gap of unknown length
* 147163: contig of 2898 bp in length
* 150061: gap of unknown length
* 150161: contig of 4631 bp in length.

```

FEATURES

```

source
  1..154791
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="7"
    /clone="RP11-595K18"
  1..5209
    /note="assembly_name:Contig10"
  5310..10071
    /note="assembly_name:Contig11"
  10172..16164
    /note="assembly_name:Contig12"
    clone_end:T7
    vector_side:left
  16265..29394
    /note="assembly_name:Contig13"
  29495..41590
    /note="assembly_name:Contig14"
  41691..54703
    /note="assembly_name:Contig15"
  54804..74381
    /note="assembly_name:Contig16"
  74482..103253
    /note="assembly_name:Contig17"
  103354..139431
    /note="assembly_name:Contig18"
  139532..140718
    /note="assembly_name:Contig15"
  clone_end:SP6
  vector_side:right
  140819..143471
    /note="assembly_name:Contig6"
  143572..147062
    /note="assembly_name:Contig7"
  147163..150060
    /note="assembly_name:Contig8"
  150161..154791

```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

ORIGIN /note="assembly_name:Contig9"

```

Query Match      8.5%; Score 205; DB 2; Length 154791;
Best Local Similarity 100.0%; Pred. No. 2.5e-107;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 312 AGCCCTAGGTGATGGCATCGCTCTCCACAGAAAGTCTTTTCCCATCTCGAAGATTG 371
Db 73243 AGCCCTAGGTGATGGCATCGCTCTCCACAGAAAGTCTTTTCCCATCTCGAAGATTG 73184
QY 372 TCTTAAGTGGCAACAACACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGCAATACCTG 431
Db 73183 TCTTAAGTGGCAACAACACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGCAATACCTG 73124
QY 432 TTTTGCCAATGCAGCACTGCAGTGTGTTAACTACACACACACCTCTTGGCCAATTACATGCT 491
Db 73123 TTTTGCCAATGCAGCACTGCAGTGTGTTAACTACACACACCTCTTGGCCAATTACATGCT 73064
QY 492 ATCAGATGAACACTCCAAAACATGT 516
Db 73063 ATCAGATGAACACTCCAAAACATGT 73039

```

RESULT 19

```

AX495102
LOCUS AX495102
DEFINITION Sequence 867 from Patent WO02059256.
ACCESSION AX495102
VERSION AX495102.1 GI:23340712
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

REFERENCE

```

1. Tuijinder, M., Telerman, A., Anson, R. and Susini, L.
Sequences involved in phenomena of tumour suppression, tumour
reversion, apoptosis and/or virus resistance and their use as
medicines

```

JOURNAL

```

Patent: WO 02059256-A 867 01-AUG-2002;

```

FEATURES

```

source
  1..213
    /organism="Homo sapiens"
    /mol_type="unassigned DNA"
    /db_xref="taxon:9606"

```

ORIGIN

```

Query Match      5.7%; Score 137; DB 6; Length 213;
Best Local Similarity 100.0%; Pred. No. 1.3e-67;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 918 GCTTGATGGAGAAACCTCGTACAAAGTGCAGAGTGTAAAAAGATGGTTCAGCTTCAA 977
Db 70 GCTTGATGGAGAAACCTCGTACAAAGTGCAGAGTGTAAAAAGATGGTTCAGCTTCAA 129
QY 978 GAGGTTCACTATCCATAGATCCTCTAATGTTCTTACACTTTTCTGTGAAACGTTTGCAAA 1037
Db 130 GAGGTTCACTATCCATAGATCCTCTAATGTTCTTACACTTTTCTGTGAAACGTTTGCAAA 189
QY 1038 TTTTACCGGTGAAAAA 1054
Db 190 TTTTACCGGTGAAAAA 206

```

RESULT 20

```

AX387329/c
LOCUS AX387329
DEFINITION Sequence 2257 from Patent WO0214500.
ACCESSION AX387329
VERSION AX387329.1 GI:19560459
KEYWORDS

```

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Escobedo, J., Garcia, P.D., Sudduth-Klinger, J., Reinhard, C.,
Randozo, F., Lamson, G., Scott, E.M., Zhang, G., Kassam, A., Pot, D. and
Labat, I.
TITLE Human genes and gene expression products
JOURNAL Patent: WO 0214500-A 2257 21-FEB-2002;
CHIRON CORPORATION (US); Hyseq Inc. (US)
FEATURES
source Location/Qualifiers
1. .292
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 4.1%; Score 100; DB 6; Length 292;
Best Local Similarity 99.3%; Pred. No. 6.2e-46;
Matches 150; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 312 AGCCCTAGTGTGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTCGAAGATTTCG 371
Dd 151 AGCCCTAGTGTGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTCGAAGATTTCG 92
QY 372 TCTTAAGTGGCAACAACACTCATAGTGTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTG 431
Dd 91 TCTTAAGTGGCAACAACACTCATAGTGTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTG 32
QY 432 TTTTGGCAATGCAGCACTGCAGTGTTTAACC 462
Dd 31 TTTTGGCAATGCAGCACTGCAGTGTTTAACC 1
RESULT 21
AC136746 195906 bp DNA linear HTG 10-MAR-2003
LOCUS Mus musculus clone RP23-151C5, WORKING DRAFT SEQUENCE, 5 unordered
DEFINITION pieces.
ACCESSION AC136746
VERSION HTG3 PHASE1; HTGS_DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
TITLE Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
JOURNAL Camarata, J., Chang, V., Chararo, B., Choepel, Y., Collymore, A.,
AUTHORS Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J.,
Mathews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, D., Oliver, J.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,
Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,
Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (07-NOV-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
TITLE
JOURNAL
AUTHORS

REFERENCE
AUTHORS 3 (bases 1 to 195906)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cook, A., Cooke, P., Dorris, L., Erickson, J., Faro, S.,
Diaz, J.S., Dodge, S., Dooley, K., Gage, D., Galagan, J., Gardyna, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, D., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Topham, K., Wu, X.,
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wyman, D.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (10-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 10, 2003 this sequence version replaced gi:28882377.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26341
Center clone name: 151C_5
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 194971 bases at least Q40
Consensus quality: 195140 bases at least Q30
Consensus quality: 195253 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 195506; sum-of-contigs
Quality coverage: 11.9 in Q20 bases; agarose-fp
Quality coverage: 11.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 104550: contig of 104650 bp in length
* 104651 104750: gap of 100 bp
* 104751 119592: contig of 14942 bp in length
* 119593 119792: gap of 100 bp
* 119793 138396: contig of 13204 bp in length
* 138397 139396: gap of 100 bp
* 139397 192208: contig of 53112 bp in length
* 192209 192308: gap of 100 bp
* 192309 195306: contig of 3598 bp in length.
* Location/Qualifiers
1. .195906
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-151C5"
/clone_lib="RPCI-23 Female Mouse BAC"
1. .104650
misc_feature

```

/note="assembly_fragment
clone_end:SP6
vector_side:left"
104751..119692
/note="assembly_fragment"
119793..138996
/note="assembly_fragment"
139097..192208
/note="assembly_fragment"
192309..195906
/note="assembly_fragment
clone_end:T7
vector_side:right"

TITLE
JOURNAL

COMMENT
Query Match          1.6%; Score 38; DB 2; Length 195906;
Best Local Similarity 100.0%; Pred. No. 2e-09;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy    10  GCGGCGGGCGGGCGGGCGGGCGGGATGGAGCGAG 47
Db    135334  GCGGCGGGCGGGCGGGCGGGCGGGATGGAGCGAG 135371

RESULT 22
AC136746/c
LOCUS
DEFINITION
Mus musculus clone RP23-151C5, WORKING DRAFT SEQUENCE, 5 unordered
pieces.
ACCESSION
AC136746
VERSION
AC136746.3 GI:28894698
KEYWORDS
HTG; HTGS PHASEL; HTGS DRAFT.
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 195906)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-151C5
Unpublished

REFERENCE
2 (bases 1 to 195906)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barn,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camrata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kellis,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Maclean,C., MacDonald,P., Major,T.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Nelson,C.H., O'Connor,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (07-NOV-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 195906)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhgalter,B., Camrata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kellis,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M. Direct Submission
Submitted (10-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 10, 2003 this sequence version replaced gi:28894698.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26341
Center clone name: 151_C_5
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 194971 bases at least Q40
Consensus quality: 195140 bases at least Q30
Consensus quality: 195253 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 195506; sum-of-contigs
Quality coverage: 11.9 in Q20 bases; agarose-fp
Quality coverage: 11.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved
* *
* 1 104550: contig of 104650 bp in length
* 104651 104750: gap of 100 bp
* 104751 119692: contig of 14942 bp in length
* 119693 119792: gap of 100 bp
* 119793 138996: contig of 19204 bp in length
* 138997 139096: gap of 100 bp
* 139097 192208: contig of 53112 bp in length
* 192209 192308: gap of 100 bp
* 192309 195906: contig of 3598 bp in length.
*
* Location/Qualifiers
*     1..195906
*         /organism="Mus musculus"
*         /mol_type="genomic DNA"
*         /db_xref="taxon:10090"
*         /clone="RP23-151C5"
*         /clone_lib="RPCI-23 Female Mouse BAC"
* misc_feature
*     1..104650
*         /note="assembly_fragment
*         clone_end:SP6
*         vector_side:left"
*     104751..119692
*         /note="assembly_fragment"
*     119793..138996
*         /note="assembly_fragment"
*     139097..192208
*         /note="assembly_fragment"
*     192309..195906
*         /note="assembly_fragment"

```

clone_end:T7
vector_side:right

ORIGIN

Query Match 1.4%; Score 34; DB 2; Length 195906;
Best Local Similarity 100.0%; Pred. No. 4.5e-07;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 TCGGTTTGAACCAAGAGATGCCCATGAATT 674
|||||
Db 99247 TCGGTTTGAACCAAGAGATGCCCATGAATT 99214

RESULT 23
AC108995/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-298L20, *** SEQUENCING IN PROGRESS
***, 5 unordered pieces.
AC108995 GI:25006701
VERSION HTG; HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
1 (bases 1 to 239456)
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Ayagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseghe,H., Lozada,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Minet,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwackeleme,O., Okwuon,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfamkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,M., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willison,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,

Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weis,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 239456)
Worley,K.C.
Direct Submission
Submitted (03-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 239456)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23195556.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPNT
Center clone name: CH230-298L20
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 22813 bases at least Q40
Consensus quality: 22529 bases at least Q30
Consensus quality: 226929 bases at least Q20
Estimated insert size: 228958; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 216639: contig of 216639 bp in length
216640 216739: gap of unknown length
216740 216728: contig of 7889 bp in length
224629 224728: gap of unknown length
224729 234582: contig of 9854 bp in length
234583 234582: gap of unknown length
234583 235702: contig of 1020 bp in length
235703 235802: gap of unknown length
235803 239456: contig of 3654 bp in length.

Location/Qualifiers
1..239456
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-298L20"
216740..218521
/note="wgs_contig"

FEATURES
source
misc_feature
216740..218521
/note="wgs_contig"


```

misc_feature 224729..226063
ORIGIN
Query Match 1.3% Score 32; DB 2; Length 239456;
Best Local Similarity 100.0%; Pred. No. 6.7e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 412 CAGAAATTGGCAATACCTCTTTTGGCAATGC 443
Db 150098 CAGAAATTGGCAATACCTCTTTTGGCAATGC 150067

RESULT 24
AC111575/c
LOCUS AC111575
DEFINITION Rattus norvegicus clone CH230-68E4, WORKING DRAFT SEQUENCE, 8
unordered pieces.
AC111575
VERSION AC111575.4 GI:24818943
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 270456)
Muzay,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,B., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Derramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregregis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensbawa,L., Louised,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olampunagoo,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabak,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,

```

Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Unpublished
2 (bases 1 to 270456)
Worley, K. C.

Direct Submission
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 270456)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 9, 2002 this sequence version replaced gi:23267991.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GMPC
Center clone name: CH230-68E4

----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 248347 bases at least Q40
Consensus quality: 250800 bases at least Q30
Consensus quality: 252288 bases at least Q20
Estimated insert size: 248995; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

	1	45421:	contig of 45421 bp in length
	45422	45521:	gap of unknown length
	45522	49934:	contig of 4413 bp in length
	49935	50034:	gap of unknown length
	50035	253199:	contig of 203165 bp in length
	253200	253299:	gap of unknown length
	253300	254387:	contig of 1088 bp in length
	254388	254487:	gap of unknown length
	254488	255765:	contig of 1278 bp in length
	255766	25865:	gap of unknown length
	25866	257214:	contig of 1349 bp in length
	257215	258888:	contig of 1574 bp in length
	258889	259888:	gap of unknown length
	258989	270456:	contig of 11468 bp in length.

Location/Qualifiers
1. :270456

FEATURES
source

```

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-68E4"
1..1277
/notes="wgs_end_extension
clone_end:Sp6"
8765..9559
/notes="clone_boundary
clone_end:Sp6
site:EcoRI
end_sequence:BH351856"
40746..40960
/notes="clone_boundary
clone_end:T7
site:EcoRI
end_sequence:BH351854"
45522..46958
/notes="wgs_end_extension
clone_end:T7"
50035..51869
/notes="wgs_end_extension
clone_end:T7"

ORIGIN
Query Match 1.3%; Score 32; DB 2; Length 270456;
Best Local Similarity 100.0%; Pred. No. 6.8e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 412 CAGAAATTGGGCATACCTGTTTGGCCAATGC 443
Db 76856 CAGAAATTGGGCATACCTGTTTGGCCAATGC 76825

RESULT 25
AX327423/c
LOCUS AX327423 637 bp DNA linear PAT 07-JAN-2002
DEFINITION Sequence 51 from Patent WO0183524.
ACCESSION AX327423
VERSION AX327423.1 GI:18097904
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS Lal, P., Yue, H., Tang, Y.T., Lu, D.A., Azimzai, Y., Au-Young, J.,
Hillman, J.L., Baughn, M.R., Yao, M.G., Burford, N., Batra, S. and
Policky, J.J.
TITLE Rna metabolism proteins
JOURNAL Patent: WO 0183524-A 51 08-NOV-2001;
Incyte Genomics, Inc. (US)
FEATURES
source
1..637
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/notes="Incyte ID No: 1647264CB1"

ORIGIN
Query Match 1.2%; Score 30; DB 6; Length 637;
Best Local Similarity 100.0%; Pred. No. 7.1e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
Db 37 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 8

RESULT 26
AX174745
LOCUS AX174745 733 bp DNA linear PAT 03-JUL-2001
DEFINITION Sequence 39 from Patent WO0142285.

```

```

AX174745
VERSION AX174745.1 GI:14598268
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS Yue, H., Tang, Y.T., Lal, P., Burford, N., Azimzai, Y., Patterson, C.,
Baughn, M.R., Lu, D.A., Shah, P. and Au-Young, J.
TITLE Extracellular matrix and cell adhesion proteins as well as genes
encoding them
JOURNAL Patent: WO 0142285-A 39 14-JUN-2001;
Incyte Genomics, Inc. (US)
FEATURES
source
1..733
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/notes="Incyte ID No: 3683905CB1"

ORIGIN
Query Match 1.2%; Score 30; DB 6; Length 733;
Best Local Similarity 100.0%; Pred. No. 7.1e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
Db 696 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 725

RESULT 27
AX078266/c
LOCUS AX078266 823 bp DNA linear PAT 22-FEB-2001
DEFINITION Sequence 70 from Patent WO0107471.
ACCESSION AX078266
VERSION AX078266.1 GI:13157957
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS Hillman, J.L., Lal, P., Tang, Y.T., Yue, H., Au-Young, J., Bandman, O.,
Azimzai, Y., Yang, J., Lu, D.A., Baughn, M.R., Patterson, C. and Shah, P.
TITLE Cell cycle and proliferation proteins
JOURNAL Patent: WO 0107471-A 70 01-FEB-2001;
Incyte Genomics, Inc. (US)
FEATURES
source
1..823
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/notes="Incyte ID No: 058336CB1"

ORIGIN
Query Match 1.2%; Score 30; DB 6; Length 823;
Best Local Similarity 100.0%; Pred. No. 7.2e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
Db 45 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 16

RESULT 28
AX073683
LOCUS AX073683 972 bp DNA linear PAT 06-FEB-2001
DEFINITION Sequence 17 from Patent WO0104264.
ACCESSION AX073683
VERSION AX073683.1 GI:12710104
KEYWORDS
SOURCE Homo sapiens (human)

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Jones, K.A., Volkmut, W. and Walker, M.G.
TITLE Atherosclerosis-associated genes
JOURNAL Patent: WO 0104264-A 17 18-JAN-2001;
Incyte Genomics, Inc. (US)
FEATURES
source
1..972
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 1.2%; Score 30; DB 6; Length 972;
Best Local Similarity 100.0%; Pred. No. 7.2e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2383 CCCTTAGTGAGGGTTAATTTAGCTTGCA 2412
|||||
Db 936 CCCTTAGTGAGGGTTAATTTAGCTTGCA 965
|||||
RESULT 29
AX135572/c 1106 bp DNA linear PAT 29-MAY-2001
LOCUS
DEFINITION Sequence 69 from Patent WO0132888.
ACCESSION AX135572
VERSION AX135572.1 GI:14271842
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Tang, Y.T., Yue, H., Hillman, J.L., Lal, P., Bandman, O., Patterson, C.,
Shih, L.L., Azimzai, Y., Lu, D.A. and Baughn, M.R.
TITLE Human transferase molecules
JOURNAL Patent: WO 0132888-A 69 10-MAY-2001;
Incyte Genomics, Inc. (US)
FEATURES
source
1..1106
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 2448909CB1"
ORIGIN
Query Match 1.2%; Score 30; DB 6; Length 1106;
Best Local Similarity 100.0%; Pred. No. 7.3e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2383 CCCTTAGTGAGGGTTAATTTAGCTTGCA 2412
|||||
Db 38 CCCTTAGTGAGGGTTAATTTAGCTTGCA 9
|||||
RESULT 30
AX135583 1163 bp DNA linear PAT 29-MAY-2001
LOCUS
DEFINITION Sequence 80 from Patent WO0132888.
ACCESSION AX135583
VERSION AX135583.1 GI:14271853
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Tang, Y.T., Yue, H., Hillman, J.L., Lal, P., Bandman, O., Patterson, C.,
Shih, L.L., Azimzai, Y., Lu, D.A. and Baughn, M.R.
TITLE Human transferase molecules

JOURNAL Patent: WO 0132888-A 80 10-MAY-2001;
Incyte Genomics, Inc. (US)
FEATURES
source
1..1163
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 4071304CB1"
ORIGIN
Query Match 1.2%; Score 30; DB 6; Length 1163;
Best Local Similarity 100.0%; Pred. No. 7.3e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2383 CCCTTAGTGAGGGTTAATTTAGCTTGCA 2412
|||||
Db 1127 CCCTTAGTGAGGGTTAATTTAGCTTGCA 1156
|||||
RESULT 31
AF121858 1487 bp mRNA linear PRI 14-SEP-2001
LOCUS
DEFINITION Homo sapiens sorting nexin 8 (SNX8) mRNA, complete cds.
ACCESSION AF121858
VERSION AF121858.1 GI:4689255
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1487)
AUTHORS Teasdale, R.D., Locci, D., Houghton, F., Karlsson, L. and Gleeson, P.A.
TITLE A large family of endosome-localized proteins related to sorting nexin 1
JOURNAL Biochem. J. 358 (Pt 1), 7-16 (2001)
MEDLINE 21378165
PubMed 11485546
REFERENCE 2 (bases 1 to 1487)
AUTHORS Teasdale, R.D., Gleeson, P.A. and Karlsson, L.
TITLE Identification of eleven novel human sorting nexin molecules. A sub-group of the sorting nexin family is associated with the early endosomes
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1487)
AUTHORS Teasdale, R.D., Gleeson, P.A. and Karlsson, L.
TITLE Direct Submission
JOURNAL Submitted (22-JAN-1999) The R.W. Johnson Pharmaceutical Research Institute, 3535 General Atomics Court, San Diego, CA 92121, USA
FEATURES
source
1..1487
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..1487
/gene="SNX8"
11..1408
/gene="SNX8"
/codon_start=1
/product="sorting nexin 8"
/protein_id="AAD27831.1"
/db_xref="GI:4689256"
/translation="MTGRAMDPLPAAAVGAAAEADPPADSLPTPOAIBPQAI
VQVQFAPSRMQPQGNFLLSHTQLIARDTVQVELIPEKGLFLKHVEYESSQRF
KSSVRRYNDVFWFQEMLLHKFVPMVMPALPKMLGADREFIARRRLKRFVNLVA
RHPLFSDVLVLFLSPGSDVONKLKESACVGVDEFNCKLAKDFADIAQAF
AISELILNIYNSEHKLDRARERIASRAIDNAADLLIFGKELSAIGSDTTPFPAAL
NSSTWGSGLKALGSLVEFALLADKAAQGGQEEENVVVKLNFLDLQLSQYKDLCEH
EKGVLHKQRLHKYSLMKRQMSATQNPRESVEQLSEIVQEENAIQTMEIRNYF
SLYCLHQTOLIHVLPILTSHILRAFNVSQIQGHKEMSKVWMDLRPKLSCLFAGPHST
LTPCSPSPEDGLCPH"
ORIGIN
Query Match 1.2%; Score 30; DB 9; Length 1487;

Best Local Similarity 100.0%; Pred. No. 7.4e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412
|||||
Db 1451 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 1480

RESULT 32
AX061211
LOCUS
DEFINITION Sequence 58 from Patent WO0078953.
ACCESSION AX061211
VERSION AX061211.1 GI:12406347
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS
TITLE Human transport proteins
JOURNAL Patent: WO 0078953-A 58 28-DEC-2000;
INCYTE Genomics, Inc. (US)
FEATURES
source
1. .1491
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 1.2%; Score 30; DB 6; Length 1491;
Best Local Similarity 100.0%; Pred. No. 7.4e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412
|||||
Db 1452 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 1481

RESULT 33
AX073668
LOCUS
DEFINITION Sequence 2 from Patent WO0104264.
ACCESSION AX073668
VERSION AX073668.1 GI:12710089
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS
TITLE Atherosclerosis-associated genes
JOURNAL Patent: WO 0104264-A 2 18-JAN-2001;
INCYTE Genomics, Inc. (US)
FEATURES
source
1. .1702
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 1.2%; Score 30; DB 6; Length 1702;
Best Local Similarity 100.0%; Pred. No. 7.5e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412
|||||
Db 1665 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 1694

RESULT 34
AX081430
LOCUS
DEFINITION Sequence 14 from Patent WO0109178.
ACCESSION AX081430
VERSION AX081430.1 GI:13170253
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS
TITLE Human chaperone proteins
JOURNAL Patent: WO 0109178-A 14 08-FEB-2001;
INCYTE Genomics, Inc. (US)
FEATURES
source
1. .1950
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 1742550CB1"

ORIGIN
Query Match 1.2%; Score 30; DB 6; Length 1950;
Best Local Similarity 100.0%; Pred. No. 7.5e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412
|||||
Db 1915 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 1944

RESULT 35
AX077793/c
LOCUS
DEFINITION Sequence 24 from Patent WO0107628.
ACCESSION AX077793
VERSION AX077793.1 GI:12157660
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS
TITLE Human synthetase
JOURNAL Patent: WO 0107628-A 24 01-FEB-2001;
INCYTE Genomics, Inc. (US)
FEATURES
source
1. .1954
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 3231586CB1"

ORIGIN
Query Match 1.2%; Score 30; DB 6; Length 1954;
Best Local Similarity 100.0%; Pred. No. 7.5e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412
|||||
Db 48 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 19

RESULT 36
BD260129/c
LOCUS
DEFINITION Matrix-remodeling genes.
ACCESSION BD260129

```

VERSION      BD260129.1  GI:33069899
KEYWORDS     JP 2002527054-A/20
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE        Walker M.G., Volkmut, W. and Klingler, T.M.
JOURNAL      Matrix-remodeling genes
Patent: JP 2002527054-A 20 27-AUG-2002;
INCYTE PHARMACEUTICALS INC
COMMENT      OS Homo sapiens (human)
PN JP 2002527054-A/20
PD 27-AUG-2002
PF 06-OCT-1999 JP 2000575891
PR 09-OCT-1998 US 09/169289
PI MICHAEL G WALKER, WAYNE VOLKUT, TOD M KLINGLER PC
C12N15/09, A61K38/00, A61K38/28, A61K48/46, A61P1/04, PC
A61P3/10,
PC A61P9/10, A61P17/02, A61P19/02, A61P35/00, C07K14/47, C07K16/18, PC
C12N1/15,
PC C12N1/19, C12N1/21, C12N5/10, C12Q1/68, G01N33/15, G01N33/50, G01N33/53,
G01N33/54, G01N33/76, C12N15/00, C12N5/00, A61K37/02,
PC G01N33/53, G01N33/54,
PC A61K37/26,
PC A61K37/54,
CC Incyte ID NO.: 3948614CB1
FH Key Location/Qualifiers
FT source 1..1987
FT /organism="Homo sapiens (human)".

FEATURES             source
    source            1..1987
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"

ORIGIN
Query Match          1.2%; Score 30; DB 6; Length 1987;
Best Local Similarity 100.0%; Pred. No. 7.6e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAATTTAGCTTGCA 2412
Db 39 CCCTTTAGTGAGGGTTAATTTAGCTTGCA 10

Search completed: August 17, 2004, 04:45:21
Job time : 14910 secs

A61P3/00,
PC A61P11/06, A61P29/00, A61P35/00, A61P37/02, A61P37/08, C07K14/47,
PC C07K16/18,
PC C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12P21/02, C12P21/08 PC
, C12Q1/68, G01N33/15,
PC G01N33/50, G01N33/53, G01N33/566, C12N15/00, C12N5/00, A61K37/02 CC
Incyte ID No:1749964CB1
FH Key Location/Qualifiers
FT source 1..3465
FT /organism="Homo sapiens (human)".

FEATURES             source
    source            1..3465
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"

ORIGIN
Query Match          1.2%; Score 30; DB 6; Length 3465;
Best Local Similarity 100.0%; Pred. No. 7.8e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAATTTAGCTTGCA 2412
Db 38 CCCTTTAGTGAGGGTTAATTTAGCTTGCA 9

Search completed: August 17, 2004, 04:45:21
Job time : 14910 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 16, 2004, 20:20:52 ; Search time 1355 Seconds
(without alignments)
7562.106 Million cell updates/sec

Title: US-10-049-745-31

Perfect score: 2412

Sequence: 1 gggctgtgcgccggcggc.....agggtaatttttagcttga 2412

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 30

Total number of hits satisfying chosen parameters: 69

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : N_Geneseq_29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002s:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2412	100.0	2412	4	Aaf81717 Human pro
2	2321	96.2	4659	6	Abk86140 CDNA enco
3	2287	94.8	2347	4	Aah19322 Human ubi
4	2287	94.8	2347	7	Abx93880 Human CDN
5	2265	93.9	3660	6	Aah17633 Human CDN
6	2210	91.6	4483	6	Abz11508 Human pol
7	1475	61.2	1679	6	Ab139772 Human NS
8	722	29.9	16682	4	Aak79798 Human imm
9	722	29.9	16682	4	Aak79798 Human imm
10	722	29.9	21436	4	Aak79799 Human imm
11	722	29.9	21436	4	Aak79799 Human imm
12	516	21.4	1122	4	AAS41077 Human CDNA enco
13	424	17.6	577	4	AAS41077 Human CDNA enco
14	411	17.0	411	4	AAL192171 Human pol
15	357	14.8	722	6	ABK30453 Human G-p
16	333	13.8	333	4	AAI22307 Probe #12
17	333	13.8	333	4	ABK30453 Human G-p
18	333	13.8	333	4	ABK30453 Human G-p
19	333	13.8	333	4	ABK30453 Human G-p
20	333	13.8	333	4	ABK30453 Human G-p
21	333	13.8	333	5	AAI07994 Human bre
22	333	13.8	333	6	ABK30453 Human G-p
23	314	13.0	498	4	AAI13071 Probe #30
24	314	13.0	498	4	ABK30453 Human G-p

C	24	314	13.0	498	4	ABA44316
C	25	314	13.0	498	4	AAK28495
C	26	314	13.0	498	5	AAI02976
C	27	314	13.0	498	6	ABK30008
C	28	232	9.6	521	4	AAH05659
C	29	205	8.5	457	6	ABK30444
C	30	205	8.5	489	4	AAI15006
C	31	205	8.5	489	4	ABA56739
C	32	205	8.5	489	4	AAI36351
C	33	205	8.5	489	4	ABA46202
C	34	205	8.5	489	4	AAK30399
C	35	205	8.5	489	4	ABK30038
C	36	205	8.5	489	5	AAI04774
C	37	205	8.5	489	6	ABK30038
C	38	198	8.2	198	4	AAI24193
C	39	198	8.2	198	4	ABA69312
C	40	198	8.2	198	4	AAI49479
C	41	198	8.2	198	4	ABA51313
C	42	198	8.2	198	4	AAK43414
C	43	198	8.2	198	4	ABK30038
C	44	198	8.2	198	4	ABK30038
C	45	198	8.2	198	5	AAI09758
C	46	137	5.7	213	6	ABK30038
C	47	100	4.1	232	6	ABK30038
C	48	85	3.5	657	4	AAK30038
C	49	30	1.2	637	6	AAK30038
C	50	30	1.2	733	4	AAK30038
C	51	30	1.2	823	4	AAK30038
C	52	30	1.2	847	7	ACA04592
C	53	30	1.2	960	3	AAK30038
C	54	30	1.2	972	4	AAK30038
C	55	30	1.2	1106	5	AAH23827
C	56	30	1.2	1112	6	AAH23827
C	57	30	1.2	1163	5	AAH23827
C	58	30	1.2	1273	8	ADA11069
C	59	30	1.2	1273	10	ADE77026
C	60	30	1.2	1491	4	AAK30038
C	61	30	1.2	1607	3	AAK30038
C	62	30	1.2	1702	4	AAK30038
C	63	30	1.2	1833	6	AAK30038
C	64	30	1.2	1926	9	ADK47516
C	65	30	1.2	1950	4	AAK30038
C	66	30	1.2	1954	4	AAK30038
C	67	30	1.2	1987	3	AAK30038
C	68	30	1.2	1987	6	AAK30038
C	69	30	1.2	3465	3	AAA08038

ALIGNMENTS

RESULT 1

AAAF81717

ID AAF81717 standard; cDNA; 2412 BP.

XX AAF81717;

AC AAF81717;

XX AAF81717;

DT 12-JUN-2001 (first entry)

XX Human protease and protease inhibitor PPTM-4 encoding cDNA.

XX Human; protease; protease inhibitor; protease and protease inhibitor;

PPIM; identification; diagnosis; anti-human immunodeficiency virus; HIV;

antidiabetic; immunostimulant; immunomodulator; antiinflammatory;

antithyroid; immunosuppressive; nephrotropic; antitumor; thyromimetic;

cytostatic; antibacterial; fungicide; protozoacide; antiarteriosclerotic;

antiatherosclerotic; antiparasitic; virucide; hepatotropic; gene therapy;

autoimmune disorder; inflammatory disease; AIDS; Chediak-Higashi syndrome;

severe combined immunodeficiency disease; SCID; Chediak-Higashi syndrome;

Cushing's disease; Addison's disease; autoimmune thyroiditis; gout;

Crohn's disease; diabetes mellitus; Good pasture's syndrome; infection;

Grave's disease; Hashimoto's thyroiditis; Sjogren's syndrome; cancer;

Werner's syndrome; cell proliferative disorder; arteriosclerosis;

CC	infarction, etc.; autoimmune/inflammatory disorders such as acquired
CC	immunodeficiency syndrome (AIDS), asthma, Grave's disease, etc.; cell
CC	proliferative disorders such as hepatitis, psoriasis, leukaemia etc;
CC	developmental disorders such as Cushing's syndrome; epithelial disorders
CC	such as dermatitis, scabies, eczema, etc.; neurological disorders such as
CC	Parkinson's disease, dementia, Alzheimer's disease, Huntington's disease,
CC	multiple sclerosis, etc; or reproductive disorders, such as infertility,
CC	CC impotence. The present sequence represents the cDNA encoding the human
CC	PRTS7 protein of the invention
XX	
SQ	Sequence 4659 BP; 1217 A; 1295 C; 1200 G; 947 T; 0 U; 0 Other;
	Query Match 96.2%; Score 2321; DB 6; Length 4659;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 2371; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	1 GGGCTGTGTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 60
DB	
QY	83 GGGCTGTGTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 142
DB	
QY	61 CAGAGTTGAACAATGACCATAGTTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTAT 120
DB	
QY	143 CAGAGTTGAACAATGACCATAGTTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTAT 202
DB	
QY	121 CAGAAATCAGCCTGCGCAGCTCCGAGGCAGTCTCACCTGAGACATGCGTCAAGTCTTGCC 180
DB	
QY	203 CAGAAATCAGCCTGCGCAGCTCCGAGGCAGTCTCACCTGAGACATGCGTCAAGTCTTGCC 262
DB	
QY	181 AGCTGGGGTGTGTGCTTCAATCGAATCATGTGTCAAAATCACACATTTCTTTAGGACCA 240
DB	
QY	263 AGCTGGGGTGTGTGCTTCAATCGAATCATGTGTCAAAATCACACATTTCTTTAGGACCA 322
DB	
QY	241 GTACCTGCTGCTGAGTTTATTCGAGTTTCATCTGTACCTGATAAATCAAAACCAATCACCA 300
DB	
QY	323 GTACCTGCTGCTGAGTTTATTCGAGTTTCATCTGTACCTGATAAATCAAAACCAATCACCA 382
DB	
QY	301 CAAAAGGATCAAGCCCTAGGTGATGGCATCGTCTCTCCACAGAAAGTCTTTTCCCATCT 360
DB	
QY	383 CAAAAGGATCAAGCCCTAGGTGATGGCATCGTCTCTCCACAGAAAGTCTTTTCCCATCT 442
DB	
QY	361 GAGAAGATTTCCTTTAACTGCGCACAAACTCATAGAGTTGAGCTGGGCTCCAGAAATTG 420
DB	
QY	443 GAGAAGATTTCCTTTAACTGCGCACAAACTCATAGAGTTGAGCTGGGCTCCAGAAATTG 502
DB	
QY	421 GGCAATACCTGTTTGGCCAAATGCGACCTGCGAGTGTTTAACTCACACCACTCTTGCC 480
DB	
QY	503 GGCAATACCTGTTTGGCCAAATGCGACCTGCGAGTGTTTAACTCACACCACTCTTGCC 562
DB	
QY	481 AATTACATGCTATACATGAAACCTCCAAAACATGTCATGCAAGAGCTTTTGTATGATG 540
DB	
QY	563 AATTACATGCTATACATGAAACCTCCAAAACATGTCATGCAAGAGCTTTTGTATGATG 622
DB	
QY	541 TGTACAATGCGAAGCACATATTACCAGGCACCTCAGTAACTCTCTGGGCAGCTTATTAACCA 600
DB	
QY	623 TGTACAATGCGAAGCACATATTACCAGGCACCTCAGTAACTCTCTGGGCAGCTTATTAACCA 682
DB	
QY	601 ATGTTTGTTCATCAATGAGATGCGGGGTATAGCTAGGCACCTCCGTTTGGAAACCAAGAA 660
DB	
QY	683 ATGTTTGTTCATCAATGAGATGCGGGGTATAGCTAGGCACCTCCGTTTGGAAACCAAGAA 742
DB	
QY	661 GATGCCCATGAATTCCTTCAATPACACTGTTGATGCTATGCAGAAAGCATGCTTGAATGCG 720
DB	
QY	743 GATGCCCATGAATTCCTTCAATPACACTGTTGATGCTATGCAGAAAGCATGCTTGAATGCG 802
DB	
QY	721 AGCAATAAATTAGACAGACACACCCAGGCCACCACTCTTGTTTGTTCAGATATTTGGAGGA 780
DB	
QY	803 AGCAATAAATTAGACAGACACACCCAGGCCACCACTCTTGTTTGTTCAGATATTTGGAGGA 862
DB	
QY	781 TACCTTAAGATCTAGAGTCAAATGTTTAAATTTGCAAGGGCGTTTCAGATCTTTTGATCCA 840
DB	
QY	863 TACCTTAAGATCTAGAGTCAAATGTTTAAATTTGCAAGGGCGTTTCAGATCTTTTGATCCA 922
DB	
QY	841 TATCTTGAATATAACATTTGGAGATAAAGGCTGCTCAGAGTGTCAAACAGGCATTTGGAGCA 900

Db	923	TATCTTGATATAACATTGGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCATTTGAGCAG	982
QY	901	TTTGTGAAGCCGGAAACAGCTTGATGGAGAAAACCTCGTACAAGTGACGAAGTGTA AAAAG	960
Db	983	TTTGTGAAGCCGGAAACAGCTTGATGGAGAAAACCTCGTACAAGTGACGAAGTGTA AAAAG	1042
QY	961	ATGGTTTCAGAGCTTCAAGAGGTTCACTATCCATAGATCCTCTAAATGTTCTTACACTTTCT	1020
Db	1043	ATGGTTTCAGAGCTTCAAGAGGTTCACTATCCATAGATCCTCTAAATGTTCTTACACTTTCT	1102
QY	1021	CTGAAACGTTTTGCPAAATTTTACCGGTGGAAAAATTTGCTAAGGATGTGAAATACCTGAG	1080
Db	1103	CTGAAACGTTTTGCPAAATTTTACCGGTGGAAAAATTTGCTAAGGATGTGAAATACCTGAG	1162
QY	1081	TATCTTGATATTGCGGCATATATGTTCTCAACCAAGGAGAGCCAAATGCTCAAGTCTGTG	1140
Db	1163	TATCTTGATATTGCGGCATATATGTTCTCAACCAAGGAGAGCCAAATGCTCAAGTCTGTG	1222
QY	1141	TATGAGTCTGCTGCACACTCGTTTTAATTGCCATGCTGGCCATTACTTTCTGTACATA	1200
Db	1223	TATGAGTCTGCTGCACACTCGTTTTAATTGCCATGCTGGCCATTACTTTCTGTACATA	1282
QY	1201	AAAGCTAGCAATGGCTCTGGTATCAAAATGAATGACTCCATTTGATATCAACAGTATTT	1260
Db	1283	AAAGCTAGCAATGGCTCTGGTATCAAAATGAATGACTCCATTTGATATCAACAGTATTT	1342
QY	1261	AGATCGGTACTGAGTCAACAAGGCTATGTTGCTCTTTTATATCAGGTCCCATGATGAAA	1320
Db	1343	AGATCGGTACTGAGTCAACAAGGCTATGTTGCTCTTTTATATCAGGTCCCATGATGAAA	1402
QY	1321	AATGGAGTGAATTAATCATCCACCCATAGCCCGGCCAGTCCCTCTCCCGCCCGCTC	1380
Db	1403	AATGGAGTGAATTAATCATCCACCCATAGCCCGGCCAGTCCCTCTCCCGCCCGCTC	1462
QY	1381	ATCAGTTCAGCGGTTGTCACCAACAAACAGGCTGCGCCAGGCTTTATCGAACACAGCTT	1440
Db	1463	ATCAGTTCAGCGGTTGTCACCAACAAACAGGCTGCGCCAGGCTTTATCGAACACAGCTT	1522
QY	1441	CCCTCTCAATGATAAAGAAATCACTTCAATGGAATGGAACCTTGAAGACACAG	1500
Db	1523	CCCTCTCAATGATAAAGAAATCACTTCAATGGAATGGAACCTTGAAGACACAG	1582
QY	1501	CCAGCAGTTCATGTCGAGTCCCTAAGGGAATTCAGTGTCAACAGGCTAGTCCCTGTT	1560
Db	1583	CCAGCAGTTCATGTCGAGTCCCTAAGGGAATTCAGTGTCAACAGGCTAGTCCCTGTT	1642
QY	1561	AATGCTTCAGTCTCTGTCAAAACTGGTCAGTTAATAGGTCTCAGTGATCCCAGAACAT	1620
Db	1643	AATGCTTCAGTCTCTGTCAAAACTGGTCAGTTAATAGGTCTCAGTGATCCCAGAACAT	1702
QY	1621	CCTAAGAAACAAAAATTACAATCAGTATTCACAAACAGTTGCTGTTCCGACGTCTCAG	1680
Db	1703	CCTAAGAAACAAAAATTACAATCAGTATTCACAAACAGTTGCTGTTCCGACGTCTCAG	1762
QY	1681	TCTCAACCTTAACCTTCATAGTAATTTCTTGGAGAACCTTACCAAGCCGCTTCCTCTCT	1740
Db	1763	TCTCAACCTTAACCTTCATAGTAATTTCTTGGAGAACCTTACCAAGCCGCTTCCTCTCT	1822
QY	1741	ACCAATTACCAATTTCTGCACTACCTCGAAACGATCTACAGTGTCAAGTTCTTAGT	1800
Db	1823	ACCAATTACCAATTTCTGCACTACCTCGAAACGATCTACAGTGTCAAGTTCTTAGT	1882
QY	1801	AAAGTAAACAAAACGATCCCGGAGTGAATCTGTCTCCAGCCGCTGATGATGCAAA	1860
Db	1883	AAAGTAAACAAAACGATCCCGGAGTGAATCTGTCTCCAGCCGCTGATGATGCAAA	1942
QY	1861	TCCAAGCTGAATCCCAAGCGTGGTCCCTATGCGGCCAGTCCCTCTGAGGACTCTGAC	1920
Db	1943	TCCAAGCTGAATCCCAAGCGTGGTCCCTATGCGGCCAGTCCCTCTGAGGACTCTGAC	2002
QY	1921	GAGGAGTCAAAAGGGGCTGGGCAAGGAAATGGGATTTGGTACGATTTGAGCTCCCACTCT	1980

Db	2003	GAGGAGTCAAAGGGGCTCGGCAAGGAGNAATGGGATTTGGTACGATTTGAGCTCCCACTCT	2062
Qy	1981	CCGGCCAAAGATGCCGAAGATGAGGAGGCCACTCCGACAGGCTTCAAGAACCCATGACC	2040
Db	2063	CCGGCCAAAGATGCCGAAGATGAGGAGGCCACTCCGACAGGCTTCAAGAACCCATGACC	2122
Qy	2041	CTAAACGGTGTCTAATAGTGTGACAGCGACAGTGACCCGAAGAAAAACGGCGTAGCGCCT	2100
Db	2123	CTAAACGGTGTCTAATAGTGTGACAGCGACAGTGACCCGAAGAAAAACGGCGTAGCGCCT	2182
Qy	2101	GATGGTGCACGTCTGCCAAGGCCAGCTGCCCTGCACTCAGAAATCCCTTGTCTAAGGCA	2160
Db	2183	GATGGTGCACGTCTGCCAAGGCCAGCTGCCCTGCACTCAGAAATCCCTTGTCTAAGGCA	2242
Qy	2161	AACGGTCTTCTCGAAAGTTGATGCCTGCTCTCTTTGTGTCTCTCCAGAAAGACAAAATC	2220
Db	2243	AACGGTCTTCTCGAAAGTTGATGCCTGCTCTCTTTGTGTCTCTCCAGAAAGACAAAATC	2302
Qy	2221	TTAGAGACCTTCAGGCTTAGCMAAACAATGAAGGCTCGACGGATGAATGAGTGACCTT	2280
Db	2303	TTAGAGACCTTCAGGCTTAGCMAAACAATGAAGGCTCGACGGATGAATGAGTGACCTT	2362
Qy	2281	GGACGACAGAGGGGCCCTCCCGAGGACCGGACGCGAGGCTCAGCCTGCGACGCCCGCC	2340
Db	2363	GGACGACAGAGGGGCCCTCCCGAGGACCGGACGCGAGGCTCAGCCTGCGACGCCCGCC	2422
Qy	2341	GCCGAATCCCTGGAGGACCGAGATGCGGCCGC	2372
Db	2423	GCCGAATCCCTGGAGGACCGAGATGCGGCCGC	2454

RESULT 3
AAH19322
ID AAH19322 standard; cDNA; 2347 BP.

AA
AC
AAH19322;

DT 20-JUL-2001 (first entry)

DE Human ubiquitin protease 23431 coding sequence.

Human; ubiquitin protease; deubiquitinating enzyme; tuberculostatic;
 antiasthmatic; antiinflammatory; antidiarrhoeic; hepatotropic;
 gynaecological; cytostatic; antimicrobial; neuroprotective; anti-HIV;
 immunosuppressive; cardiact; antinaemic; nephrotropic; antibacterial;
 anti-thyroid; gastrointestinal; ss.

XX
OS
Homo sapiens.

AA		Key	Location/Qualifiers
FH		CDS	61..2346
FT			/+tag= a
FT			/partial
FT			/product= "Human ub
FT			/note= "No stop cod

XX
PN WO200123589-A2.

05-APR-2001.

PF 29-SEP-2000; 2000WO-US026962.

XX
PR 29-SEP-1999; 99US-00407356.

PA (MILL-) MILLENNIUM PHARM INC.

PI Kapeller-Libermann R;

DR WPI; 2001-374253/39.

DR F-PSDB; AAB82177.
XX

PT enzymes is useful for diagnosis and treatment of e.g. tuberculosis and

PT Alzheimer's disease.

PS Claim 1; Fig 1; 111pp; English.

The present sequence is the coding sequence for human ubiquitin protease 23431. The ubiquitin protease coding sequence and protein are useful as reagents or targets in assays for treatment and diagnosis of ubiquitin-mediated or -related disorders, especially disorders mediated by deubiquitinating enzymes. The protein and coding sequence are also useful for treating disorders involving the following: the spleen e.g. tuberculosis and typhoid fever, the lung such as bronchial asthma, the colon e.g. as diarrhoea and dysentery, the liver e.g. jaundice and cholestasis, the uterus and endometrium e.g. endometriosis, the brain e.g. acute meningitis and Alzheimer disease, T-cells including transplant rejection and autoimmune diseases such as systemic lupus erythematosus, diseases of the skin such as malignant melanoma, the bone marrow e.g. B- and T-lymphoid leukaemias, the heart including myocardial infarction, red cells e.g. anaemia, the thymus e.g. Hodgkin disease, B-cells e.g. peripheral B-cell neoplasms, the kidney e.g. polycystic kidney disease, the breast including periductal mastitis, the testis and epididymis e.g. syphilis, the prostate e.g. nodular hyperplasia, the thyroid, e.g. hyperthyroidism, the skeletal muscle e.g. rhabdomyosarcoma, the pancreas e.g. ectopic pancreas, reduced platelet number e.g. HIV (human immunodeficiency virus)-associated thrombocytopenia and disorders involving precursor T-cell neoplasms including precursor T lymphoblastic leukaemia/lymphoma

Sequence 2347 BP; 647 A; 606 C; 528 G; 566 T; 0 U; 0 Other;

Query Match 94.8%; Score 2287; DB 4; Length 2347;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 2337; Conservative 0; Mismatches

QY 22 GCGGCGCCGAGGGGATGGAGCGGCGCGGTCAGAGTTGAACAATGACCAT 81

Db
10 GCGGCGCCGAGCGGGCATGGAGCGAGCGCGGTCAGAGTTGAACAATGACCATTA 59

82 GTTGACAAAGGTTTCATCTTCAATCCCATCAGCGGTATCACATACTACATTA

[illegible][illegible]

130

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

100

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible][illegible]

36
35
34
33
32
31
30
29
28
27
26
25
24
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1

1-2 GAGGCGGGCCGCCGCCTCCTCCACGAGTATTCCTTTCCCCCATCTGAGGAGAGATTGTCTTAAGTGG 369

362 CACACACACATAGAGTGGAGCTGGGCTCCAGAAATTGGGCAATACCCTGTTTGGCCAAT 441

DD 370 CAACAAACTCATAGAGTTGGAGCTGGGCCTCCAGAAATTGGGCAATACCTGTTTGGCCAAT 429

442 GCAGCACTGCAGTGTTAACCTACACACCACCTCTTGCCAATTACATGCTATCACATGAA 501

DB 430 GCAGCACTGCAGTGTTAACCTACACACCACCTCTTGCCAATTACATGCTATCACAATGAA 489

QY 502 CACTCCAAACATGTTCATGCAGAAAGGCTTTTGTATGATGTGTACAATGCAAGCACATAT 561

Db 490 CACTCCAAACATGTCATGCAGAGGCTTTGTATGATGTGTACAATGCAAGCACATATT 549

QY 562 ACCCAGGCACTCAGTAATCCTGGGACGTTATTAAACCAATGTTGTCAATGAGATG 621

D_b 550 ACCCAGGCACTCAGTAATCCTGGGACGTTATTAAACCAATGTTTGTCAATCAATGAGATG 609

QY	622	CGCGGTATAGCTAGGCACCTCCGTTTGGAAACCAAGAGATGCCCATGAATTCCTTCAA	681
DB	610	CGCGGTATAGCTAGGCACCTCCGTTTGGAAACCAAGAGATGCCCATGAATTCCTTCAA	669
QY	682	TACACTGTTGATGCTATGCAGAAAGCATGCTTGAATGGCAGCAATAAATATAGACGACAC	741
DB	670	TACACTGTTGATGCTATGCAGAAAGCATGCTTGAATGGCAGCAATAAATATAGACGACAC	729
QY	742	ACCGAGGCCACCACTCTTGTGTGTCAGATATTTGGAGGATACCTTAGATCTAGATGCTCAA	801
DB	730	ACCGAGGCCACCACTCTTGTGTGTCAGATATTTGGAGGATACCTTAGATCTAGATGCTCAA	789
QY	802	TGTTTAAATTGCAAGGCGTTTCAGATACCTTTTGATCCCATATCTTGATATAACATGGAG	861
DB	790	TGTTTAAATTGCAAGGCGTTTCAGATACCTTTTGATCCCATATCTTGATATAACATGGAG	849
QY	862	ATAAGSGCTGCTCAGAGTGTCACAAAGGCATTTGGAGCATTTGTGTAAGCCGGAACAGCTT	921
DB	850	ATAAGSGCTGCTCAGAGTGTCACAAAGGCATTTGGAGCATTTGTGTAAGCCGGAACAGCTT	909
QY	922	GATGGAGAAACCTGTCACAGTCGACCAAGTGTAAAGAGATGTTCCAGCTTCAAGAGG	981
DB	910	GATGGAGAAACCTGTCACAGTCGACCAAGTGTAAAGAGATGTTCCAGCTTCAAGAGG	969
QY	982	TTCACTATCCATAGATCCTCTAATGTTCTTACACTTTTCTCGAAACGTTTTCGAAATTTT	1041
DB	970	TTCACTATCCATAGATCCTCTAATGTTCTTACACTTTTCTCGAAACGTTTTCGAAATTTT	1029
QY	1042	ACCGTGGAAAAATGTCTAAGGATGTGAATAACCTGAGTATCTTGATATTTGGCCCATAT	1101
DB	1030	ACCGTGGAAAAATGTCTAAGGATGTGAATAACCTGAGTATCTTGATATTTGGCCCATAT	1089
QY	1102	ATGCTCTCAACCCACGAGAGCCAAATGTTCTAGTCTTGATGTCAGTGTGGTCCACACT	1161
DB	1090	ATGCTCTCAACCCACGAGAGCCAAATGTTCTAGTCTTGATGTCAGTGTGGTCCACACT	1149
QY	1162	GGTTTTAAATTGCCATGCTGGCCATTACTTCTGCTACATAAAAGCTAGCAATGGCTCTCG	1221
DB	1150	GGTTTTAAATTGCCATGCTGGCCATTACTTCTGCTACATAAAAGCTAGCAATGGCTCTCG	1209
QY	1222	TATCAAAATGAATGACTCCATTTGATATACAGTGATATTTAGATCGGTACTCAGCCAAACAA	1281
DB	1210	TATCAAAATGAATGACTCCATTTGATATACAGTGATATTTAGATCGGTACTCAGCCAAACAA	1269
QY	1282	GCCTATGCTCTTTTATATCAGTCCCATGATGTGAAAAATGGAGTGAATTTACTCAT	1341
DB	1270	GCCTATGCTCTTTTATATCAGTCCCATGATGTGAAAAATGGAGTGAATTTACTCAT	1329
QY	1342	CCCAACCATAGCCCGGCCAGTCTCTCCCGCCCGCTCATCATGTCAGCGGGTTGCACC	1401
DB	1330	CCCAACCATAGCCCGGCCAGTCTCTCTCCCGCCCGCTCATCATGTCAGCGGGTTGCACC	1389
QY	1402	AACAAACAGGCTGGCCAGGCTTTATCGGACCAAGCTTCCCTCTCATATGATAAAGAAT	1461
DB	1390	AACAAACAGGCTGGCCAGGCTTTATCGGACCAAGCTTCCCTCTCATATGATAAAGAAT	1449
QY	1462	CCACTCATTTAAATGGACTGGACCATTGAAGACACGCAAGCAGTTCATATGTCAGT	1521
DB	1450	CCACTCATTTAAATGGACTGGACCATTGAAGACACGCAAGCAGTTCATATGTCAGT	1509
QY	1522	CCTAACGGGAATTCAGTGTCACAGGGCTAGTCTCTTAAATGCTTCAGCTTCTGTCCAA	1581
DB	1510	CCTAACGGGAATTCAGTGTCACAGGGCTAGTCTCTTAAATGCTTCAGCTTCTGTCCAA	1569
QY	1582	AACGTGTCAGTTAATAGTCTCTAGTGATCCAGAACATCTCTAGAAACCAAAAATTACA	1641
DB	1570	AACGTGTCAGTTAATAGTCTCTAGTGATCCAGAACATCTCTAGAAACCAAAAATTACA	1629
QY	1642	ATCAGTATTCACAAACAGTTGCTGTTTCGCGCAGTGTGAGTCTCAACCTAACCTTCATAGT	1701
DB	1630	ATCAGTATTCACAAACAGTTGCTGTTTCGCGCAGTGTGAGTCTCAACCTAACCTTCATAGT	1689
QY	1702	AAATCTTTTGGAGAACCCCTACCAAGCCGTTCCCTTCTTCTACCAATACCAATCTGCGAGTA	1761

Db	1690	1749
Db	1690	AATTTCTTTGGAGAACCTTACCAAGCCGGTTCCCTCTTCTACCAATTACCAATTTCTGCAGTA
Qy	1762	CAGTCTACCTTCGAACGCATCTACGATGTCTAGTTTCTAGTTAAAGTAAACAAACCGATCCCC
Db	1750	CAGTCTACCTTCGAACGCATCTACGATGTCTAGTTTCTAGTTAAAGTAAACAAACCGATCCCC
Qy	1822	CGCAGTGAATCTCTGCTCCCGACCCCGTGATGAATGGCAAAATCCAAAGCTGAATCCACGGCTG
Db	1810	CGCAGTGAATCTCTGCTCCCGACCCCGTGATGAATGGCAAAATCCAAAGCTGAATCCACGGCTG
Qy	1882	CTGGTGGCCCTATGGCGCGGAGTCTCTGAGGACTCTTGACGAGGAGTCAAAAGGGCTGGGC
Db	1870	CTGGTGGCCCTATGGCGCGGAGTCTCTGAGGACTCTTGACGAGGAGTCAAAAGGGCTGGGC
Qy	1942	AAGCAGAAATGGGAATGGTACGATTGTGAGCTCCCACTCTCCCGGCCCAAGATGCCGAAGAT
Db	1930	AAGCAGAAATGGGAATGGTACGATTGTGAGCTCCCACTCTCCCGGCCCAAGATGCCGAAGAT
Qy	2002	GAGGAGGCCACTCCGACAGAGCTTCAAGAACCCATGACCCCTAAACCGTGTCTAATAGTGCA
Db	1990	GAGGAGGCCACTCCGACAGAGCTTCAAGAACCCATGACCCCTAAACCGTGTCTAATAGTGCA
Qy	2062	GACAGCGACAGTGCACCGAAGAAACGGCTAGCGCTGATGTCACGCTGCCAAGGC
Db	2050	GACAGCGACAGTGCACCGAAGAAACGGCTAGCGCTGATGTCACGCTGCCAAGGC
Qy	2122	CAGCCTGCCCTGCACTCAGAAAAATCCCTTTGCTTAAGGCAAAACGGTCTTCTCTGGAAAGTTG
Db	2110	CAGCCTGCCCTGCACTCAGAAAAATCCCTTTGCTTAAGGCAAAACGGTCTTCTCTGGAAAGTTG
Qy	2182	ATGCTCTCTCTTTGCTGCTCTCTCCAGAGACAAAAATCTTAGAGACTTCAGCCTTAGC
Db	2170	ATGCTCTCTCTTTGCTGCTCTCTCCAGAGACAAAAATCTTAGAGACTTCAGCCTTAGC
Qy	2242	AACAACTGAAGGCTCGACGGATGAATCGATGTCACCTGGACAGAGAGGGCCCTCCC
Db	2230	AACAACTGAAGGCTCGACGGATGAATCGATGTCACCTGGACAGAGAGGGCCCTCCC
Qy	2302	GAGGACCGGACGCGGAGCCTCAGCCTGGACGCCCGCCGCGCAATCCCTGGAGGAGC
Db	2290	GAGGACCGGACGCGGAGCCTCAGCCTGGACGCCCGCCGCGCAATCCCTGGAGGAGC
RESULT 4		
ABX93880		
ID	ABX93880	standard; cDNA, 2347 BP.
XX	AC	ABX93880;
XX	DT	10-JUN-2003 (first entry)
XX	DE	Human cDNA encoding ubiquitin protease.
XX	XX	Human; ss; gene; ubiquitin protease; spleen disorder; Hodgkin's disease;
KW	KW	lung disorder; adult respiratory distress syndrome; colon disorder;
KW	KW	inflammatory bowel disease; liver disorder; jaundice; uterine disorder;
KW	KW	endometriosis; brain disorder; Alzheimer's disease; T-cell disorder;
KW	KW	acquired immunodeficiency syndrome; AIDS; skin disorder; urticaria;
KW	KW	heart disorder; ischaemic heart disease; blood vessel disorder;
KW	KW	atherosclerosis; red blood cell disorder; anaemia; thymus disorder;
KW	KW	DiGeorge syndrome; B-cell disorder; leukaemia; kidney disorder;
KW	KW	polycystic kidney disease; glomerulonephritis; breast disorder; mastitis;
KW	KW	testicular disorder; sexually transmitted disease; thyroid disorder;
KW	KW	hypothyroidism; pancreatic disorder; pancreatitis; intestinal disorder;
KW	KW	whipple disease; tumour; cancer.
XX	OS	Homo sapiens.
XX	FX	
XX	Key	Location/Qualifiers
FT	CDS	61..2347
FT	FT	/*tag= a
FT	FT	/product= "Ubiquitin protease"

Db	1090	ATGTC	CAACCC	ACGAGAG	CCAA	TGCT	TCTAG	CTTGT	TATG	CTAG	CTTGG	TCTG	CACT	CTGG	TCC	CACT	1145												
Qy	1162	GGTTTTAA	TTC	CCAT	CTG	TG	CCCA	TAT	CT	TG	CT	CA	TAAA	AG	CTAG	CAAT	GG	CT	TGG	1221									
Db	1150	GGTTTTAA	TTC	CCAT	CTG	TG	CCCA	TAT	CT	TG	CT	CA	TAAA	AG	CTAG	CAAT	GG	CT	TGG	1209									
Qy	1222	TATCAAA	TGAAT	GAC	TCT	CA	TTC	AT	CT	TAC	CAG	TGA	TAT	TAG	AT	CG	TACT	CG	TAC	GA	CA	AA	1281						
Db	1210	TATCAAA	TGAAT	GAC	TCT	CA	TTC	AT	CT	TAC	CAG	TGA	TAT	TAG	AT	CG	TACT	CG	TAC	GA	CA	AA	1269						
Qy	1282	GCCTAT	GTG	CT	CT	T	TAT	A	TAC	AG	GT	CC	AT	GA	GT	G	AAAA	TG	GA	AG	GT	GA	AA	1341					
Db	1270	GCCTAT	GTG	CT	CT	T	TAT	A	TAC	AG	GT	CC	AT	GA	GT	G	AAAA	TG	GA	AG	GT	GA	AA	1329					
Qy	1342	CCAC	CCAT	AG	CC	CG	CG	CA	GT	CT	CT	CC	CG	CG	CG	CT	CA	T	CAG	T	CAG	CG	GG	TT	GT	CAC	1401		
Db	1330	CCAC	CCAT	AG	CC	CG	CG	CA	GT	CT	CT	CC	CG	CG	CG	CT	CA	T	CAG	T	CAG	CG	GG	TT	GT	CAC	1389		
Qy	1402	AACAA	CAG	CG	TG	CG	CAG	CT	T	TAT	CG	AC	CA	CAG	CT	T	CC	CT	CA	CA	TG	AT	AA	GA	AT	1466			
Db	1390	AACAA	CAG	CG	TG	CG	CAG	CT	T	TAT	CG	AC	CA	CAG	CT	T	CC	CT	CA	CA	TG	AT	AA	GA	AT	1449			
Qy	1462	CCAC	CT	CA	CT	TAA	TG	GA	CT	GA	CA	CT	TAA	AG	CA	CG	CA	GA	GT	CC	AT	TC	TG	CG	AG	1521			
Db	1450	CCAC	CT	CA	CT	TAA	TG	GA	CT	GA	CA	CT	TAA	AG	CA	CG	CA	GA	GT	CC	AT	TC	TG	CG	AG	1509			
Qy	1522	CCT	TA	CC	GG	AA	T	CC	AG	GT	CA	CA	GG	TG	CT	CT	T	TA	AT	CT	CT	C	TG	CT	CT	CA	1581		
Db	1510	CCT	TA	CC	GG	AA	T	CC	AG	GT	CA	CA	GG	TG	CT	CT	T	TA	AT	CT	CT	C	TG	CT	CT	CA	1569		
Qy	1582	AACT	TG	GT	CAG	T	AA	TAG	GT	CT	CT	CAG	AT	CC	CA	GA	CA	T	CC	T	AA	GA	AA	CA	AA	AA	TT	CA	1641
Db	1570	AACT	TG	GT	CAG	T	AA	TAG	GT	CT	CT	CAG	AT	CC	CA	GA	CA	T	CC	T	AA	GA	AA	CA	AA	AA	TT	CA	1629
Qy	1642	AT	CAG	TAT	T	CA	CA	AA	GT	GC	CT	TG	CC	AG	GT	CT	CAG	T	CT	CA	CT	AA	CT	AA	CT	CT	CA	GT	1701
Db	1630	AT	CAG	TAT	T	CA	CA	AA	GT	GC	CT	TG	CC	AG	GT	CT	CAG	T	CT	CA	CT	AA	CT	AA	CT	CT	CA	GT	1689
Qy	1702	AA	TT	CT	TT	TG	GA	AA	CC	T	AC	CA	GG	CG	TT	T	CC	CT	T	CT	AC	CA	TT	AC	CA	TT	CT	GA	1761
Db	1690	AA	TT	CT	TT	TG	GA	AA	CC	T	AC	CA	GG	CG	TT	T	CC	CT	T	CT	AC	CA	TT	AC	CA	TT	CT	GA	1749
Qy	1762	CAG	T	CT	AC	CT	CG	AA	CG	AT	CT	TA	CG	AT	CT	AG	TA	AG	TA	AG	TA	AG	TA	AG	TA	AG	TA	AG	1821
Db	1750	CAG	T	CT	AC	CT	CG	AA	CG	AT	CT	TA	CG	AT	CT	AG	TA	AG	TA	AG	TA	AG	TA	AG	TA	AG	TA	AG	1809
Qy	1822	CG	CAG	T	GA	AT	CT	CG	CC	AG	CG	T	GA	T	GA	T	GA	T	GA	T									

Qy	2242	AACAACTGAAAGGCTCGAGGATGAATCAGTGCACCTGGAGCAGAGAGGGCCCTCCC	2303
Db	2230	AACAAACTGAAAGGCTCGAGGATGAATCAGTGCACCTGGAGCAGAGAGGGCCCTCCC	2289
Qy	2302	GAGGACCGGACGCCGAGCCTTCAGCTGGCAGCCCCGCCGCAATCCCTGGAGGAGC	2359
Db	2290	GAGGACCGGACGCCGAGCCTTCAGCTGGCAGCCCCGCCGCAATCCCTGGAGGAGC	2347
RESULT 5			
AAH17633			
ID	AAH17633 standard; cDNA; 3660 BP.		
XX	AAH17633;		
AC			
XX	26-JUN-2001 (first entry)		
DT			
XX	Human cDNA sequence SEQ ID NO:17168.		
DE			
XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss		
KW			
XX	Homo sapiens.		
OS			
XX	EP1074617-A2.		
PN			
XX			
XX	07-FEB-2001.		
PD			
XX	28-JUL-2000; 2000EP-00116126.		
PF			
XX	29-JUL-1999; 99JP-00248036.		
PR			
XX	27-AUG-1999; 99JP-0030253.		
PR			
XX	11-JAN-2000; 2000JP-00118776.		
PR			
XX	02-MAY-2000; 2000JP-00163767.		
PR			
XX	09-JUN-2000; 2000JP-00241899.		
XX			
PA	(HELI-) HELIX RES INST.		
XX			
PI	Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
XX			
XX	WPI; 2001-318749/34.		
DR			
XX			
PT	Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.		
PT			
PS	Claim 8; SEQ ID NO 17168; 2537pp + Sequence Listing; English.		
XX			
CC	The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises:		
CC	(a) an oligo-dt primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent polynucleotides, all of which are used in the exemplification of the present invention		
XX			

[illegible]

Qy	1026	ACGTTTTCGCAATTTTACCGGTGGAATAATTCCTAAGATGTGAAATACCTCGATATCT	1081
Db	1021	ACGTTTTCGCAATTTTACCGGTGGAATAATTCCTAAGATGTGAAATACCTCGATATCT	1080
Qy	1086	TGATATTCGGCCATATATGCTCAACCCACGGAGAGCAATTCCTACGTCCTGTATGC	1145
Db	1081	TGATATTCGGCCATATATGCTCAACCCACGGAGAGCAATTCCTACGTCCTGTATGC	1140
Qy	1146	AGTCTGCTCCACACTGTTTAAATTCGCCATGCTGGCCATTCTTCTGCTACATAAAAGC	1205
Db	1141	AGTCTGCTCCACACTGTTTAAATTCGCCATGCTGGCCATTCTTCTGCTACATAAAAGC	1200
Qy	1206	TAGCAATGGCCCTCTGGTATCAAAATGAATGACTCCATTTGTAATCTACCAAGTATATAGATC	1265
Db	1201	TAGCAATGGCCCTCTGGTATCAAAATGAATGACTCCATTTGTAATCTACCAAGTATATAGATC	1260
Qy	1266	GGTACTCAGCCCAACAAGCCTATGCTCTTTTATATCAGTCCCATGATGTGAAAAATGG	1325
Db	1261	GGTACTCAGCCCAACAAGCCTATGCTCTTTTATATCAGTCCCATGATGTGAAAAATGG	1320
Qy	1326	AGGTGAACCTTACTCATCCCAACCATATGCCCGGCCAGTCTCTCCCGCCGCTCATCAG	1385
Db	1321	AGGTGAACCTTACTCATCCCAACCATATGCCCGGCCAGTCTCTCCCGCCGCTCATCAG	1380
Qy	1386	TCAGCGGGTTGTCAACCAACAAACAGGCTGCGCCAGGCTTTATCGGACCAACAGCTTCCCTC	1445
Db	1381	TCAGCGGGTTGTCAACCAACAAACAGGCTGCGCCAGGCTTTATCGGACCAACAGCTTCCCTC	1440
Qy	1446	TCACATGATAAAGATCCACCTCATTAAATGGGACTGGACCAATTGAAAGACACGCCAAG	1505
Db	1441	TCACATGATAAAGATCCACCTCATTAAATGGGACTGGACCAATTGAAAGACACGCCAAG	1500
Qy	1506	CAGTTCCATGTCAGAGTCCCTAAACGGGAATTCAGTGTCAACAGGCTAGTCCCTGTTAATGC	1565
Db	1501	CAGTTCCATGTCAGAGTCCCTAAACGGGAATTCAGTGTCAACAGGCTAGTCCCTGTTAATGC	1560
Qy	1566	TTCAGCTTCCTGCCAAAACCTGGTCAGTTAATAGTCTCAGTGTATCCCAAGACATCCTAA	1625
Db	1561	TTCAGCTTCCTGCCAAAACCTGGTCAGTTAATAGTCTCAGTGTATCCCAAGACATCCTAA	1620
Qy	1626	GAACAACAAAATTTACAATCAGTATTCAACAAGTTGCTGTTGCCAGTGTCAGTCTCA	1685
Db	1621	GAACAACAAAATTTACAATCAGTATTCAACAAGTTGCTGTTGCCAGTGTCAGTCTCA	1680
Qy	1686	ACCTAACCTTCATAGTAAATTTCTTTGGAGAACCTACCAAGCCGTTCCCTCTCTACCAT	1745
Db	1681	ACCTAACCTTCATAGTAAATTTCTTTGGAGAACCTACCAAGCCGTTCCCTCTCTACCAT	1740
Qy	1746	TACCAATTCGCAATACAGTCTACCTCGAACGCATCTACGATGTCAAGTTTCTAGTAAAGT	1805
Db	1741	TACCAATTCGCAATACAGTCTACCTCGAACGCATCTACGATGTCAAGTTTCTAGTAAAGT	1800
Qy	1806	AACAAACCGATCCCGCAGTGAAATCTGTCTCCAGCCCGTGATGAATGGCAAAATCCAA	1865
Db	1801	AACAAACCGATCCCGCAGTGAAATCTGTCTCCAGCCCGTGATGAATGGCAAAATCCAA	1860
Qy	1866	GCTGAATCTCACGCTGCTGGTCCCTATGGCGCCAGTCTCTCTGAGGACTCTGACGAGA	1925
Db	1861	GCTGAATCTCACGCTGCTGGTCCCTATGGCGCCAGTCTCTCTGAGGACTCTGACGAGA	1920
Qy	1926	GTCAAGGGGCTGGGCAAGAGAAATGGATTTGGTACGATCTGAGCTCCCACTCTCCGG	1985
Db	1921	GTCAAGGGGCTGGGCAAGAGAAATGGATTTGGTACGATCTGAGCTCCCACTCTCCGG	1980
Qy	1986	CCAAGATGCCGAAGATGAGGAGGCCATCTCCGACGAGCTTCAAGAACCCATGACCTAAA	2045
Db	1981	CCAAGATGCCGAAGATGAGGAGGCCATCTCCGACGAGCTTCAAGAACCCATGACCTAAA	2040
Qy	2046	CGGTGCTAATAGTCAGACAGGACAGTGAACCCGAAGAAAACGGCTAGCGCTGATGG	2105
Db	2041	CGGTGCTAATAGTCAGACAGGACAGTGAACCCGAAGAAAACGGCTAGCGCTGATGG	2100

Db	1019	TTTTGTAAGCCGGAAACAGCTTGTATGGAGAAAACCTCGTACAAGTCGACGAAGTGTAAAAAG	1078
Qy	961	ATGGTTCAGCTTCAAAGAGCTTCACATATCCATAGATCCTCTAATGTCTTAAATGTCTTACACTTTCT	1020
Db	1079	ATGGTTCAGCTTCAAAGAGCTTCACTATCCATAGATCCCTTAAGTGTCTTCACTTCT	1138
Qy	1021	CTGAAACGTTTTGCAATTTTACCGGTGGAAAAATTGCTAAGGATGTGAAATACCTTGAG	1080
Db	1139	CTGAAACGTTTTGCAATTTTACCGGTGGAAAAATTGCTAAGGATGTGAAATACCTTGAG	1198
Qy	1081	TATCTTGATATTCGGCCATATATCTCTCAACCCAAACGAGAGCCAAATGTCTAGCTTG	1140
Db	1199	TATCTTGATATTCGGCCATATATCTCTCAACCCAAACGAGAGCCAAATGTCTAGCTTG	1258
Qy	1141	TATCAGTGTGTTCCACACTGGTTTTAAATGCCATGCTGCCAATTACTTCTGCTACATA	1200
Db	1259	TATCAGTGTGTTCCACACTGGTTTTAAATGCCATGCTGCCAATTACTTCTGCTACATA	1318
Qy	1201	AAAGCTAGCAATGGCCCTCTGGTATCAAAATGAATGACTTCCATGTGATCTACAGTGATTT	1260
Db	1319	AAAGCTAGCAATGGCCCTCTGGTATCAAAATGAATGACTTCCATGTGATCTACAGTGATTT	1378
Qy	1261	AGATCGGTACTCAGCCAAACAGCTATGTCTCTTTTATATCAGTGTCCATGATGTGAAA	1320
Db	1379	AGATCGGTACTCAGCCAAACAGCTATGTCTCTTTTATATCAGTGTCCATGATGTGAAA	1438
Qy	1321	AATGGAGGTGAACCTTACTCATCCCAACCATAGCCCGCCAGTCTCTCTCCGCCCCGCTC	1380
Db	1439	AATGGAGGTGAACCTTACTCATCCCAACCATAGCCCGCCAGTCTCTCTCCGCCCCGCTC	1498
Qy	1381	ATCAGTCAGCGGTTGTCAACAACACAGGCTGCGCAGGCTTTATCGACACACAGCTT	1440
Db	1499	ATCAGTCAGCGGTTGTCAACAACACAGGCTGCGCAGGCTTTATCGACACACAGCTT	1558
Qy	1441	CCCTCTCACATGATAAGAAATCCACCTCACTTAAATGGGACTGGACGATTTGAAAGACAG	1500
Db	1559	CCCTCTCACATGATAAGAAATCCACCTCACTTAAATGGGACTGGACGATTTGAAAGACAG	1618
Qy	1501	CCAAGCAGTTCCTATGTCCAGTCTCTAACGGGAAATTCAGTGTCAACAGGCTAGTCCCTGT	1560
Db	1619	CCAAGCAGTTCCTATGTCCAGTCTCTAACGGGAAATTCAGTGTCAACAGGCTAGTCCCTGT	1678
Qy	1561	AATGCTTCAGCTTCTGTCCAAAACCTGCTCAGTTAATAGTCTCTCAGTGATGCCAGAACAT	1620
Db	1679	AATGCTTCAGCTTCTGTCCAAAACCTGCTCAGTTAATAGTCTCTCAGTGATGCCAGAACAT	1738
Qy	1621	CCTAAGAAACAAAATATCAATCAGTATTCACAAAGTTGGCTGCTTTCGCCAGTGTGAG	1680
Db	1739	CCTAAGAAACAAAATATCAATCAGTATTCACAAAGTTGGCTGCTTTCGCCAGTGTGAG	1798
Qy	1681	TCTCAACCTAACCTTCATAGTAATTTCTTGGAGAACCTTACCAAGCCGTTCCCTCTTCT	1740
Db	1799	TCTCAACCTAACCTTCATAGTAATTTCTTGGAGAACCTTACCAAGCCGTTCCCTCTTCT	1858
Qy	1741	ACCATTAACAAATCTCGAGTACAGTCTAGCTCGAACGCAATCTACAGATGTCACTTCTAGT	1800
Db	1859	ACCATTAACAAATCTCGAGTACAGTCTAGCTCGAACGCAATCTACAGATGTCACTTCTAGT	1918
Qy	1801	AAAGTAAACAAACCGATCCCCCGCAGTGAATCTGTCTCCAGCCGCGTGAATGGCAAA	1860
Db	1919	AAAGTAAACAAACCGATCCCCCGCAGTGAATCTGTCTCCAGCCGCGTGAATGGCAAA	1978
Qy	1861	TCCAAGCTGAATCCAGCGTGTGGTGCCCTATGGCGCCGAGTCTCTCAGGACTCTGAC	1920
Db	1979	TCCAAGCTGAATCCAGCGTGTGGTGCCCTATGGCGCCGAGTCTCTCAGGACTCTGAC	2038
Qy	1921	GAGGAGTCAAGGGGCTGGGCAAGAGAAATGGGATTTGGTACGAATGTGAGCTCCCACTCT	1980
Db	2039	GAGGAGTCAAGGGGCTGGGCAAGAGAAATGGGATTTGGTACGAATGTGAGCTCCCACTCT	2098
Qy	1981	CCCGGCCAAGATCCGGAAGATGAGGAGGCAATCCGCAAGCTTCAAGAACCCCATGACC	2040

Db	2099	CCCGGCCAAGATGCCGAAGATGAGGAGGCCACCTCCGCACGAGCTTCAAGAACCCATGACC	2155
Qy	2041	CTAAACGGTGTCTAATAGTGCAGACAGCGACAGTCAACCGGAAAGAAAACGGCTAGCGCCT	2100
Db	2159	CTAAACGGTGTCTAATAGTGCAGACAGCGACAGTCAACCGGAAAGAAAACGGCTAGCGCCT	2218
Qy	2101	GATGGTGCAGCTGCCAAGGCCAGCCCTGCCTGCACCTCAGAAAAATCCCTTTGCTAAGGCA	2160
Db	2219	GATGGTGCAGCTGCCAAGGCCAGCCCTGCCTGCACCTCAGAAAAATCCCTTTGCTAAGGCA	2278
Qy	2161	AACGGTCTTCTTGGAAAGTTGATGCCCTGCTCCTTTGCTGTCTCTCCAGAAAGACAAAATC	2220
Db	2279	AACGGTCTTCTTGGAAAGTTGATGCCCTGCTCCTTTGCTGTCTCTCCAGAAAGACAAAATC	2338
Qy	2221	TTAGAGACCTTCAGGCTTACGACAACTGAAAGGCTGACGGATGAATGATGTGACCT	2280
Db	2339	TTAGAGACCTTCAGGCTTACGACAACTGAAAGGCTGACGGATGAATGATGTGACCT	2398
Qy	2281	GGAGCAGAGAGGGGCCCTCCCGAGGACCGCGACGCGAGCCTCAGCCTGCGACCCCGGCC	2340
Db	2399	GGAGCAGAGAGGGGCCCTCCCGAGGACCGCGACGCGAGCCTCAGCCTGCGACCCCGGCC	2458
Qy	2341	GCCGAATCCCTGGAGGAGCCAGATGGCGCGC	2372
Db	2459	GCCGAATCCCTGGAGGAGCCAGATGGCGCGC	2490
RESULT 7			
ABL39772			
ID	ABL39772 standard; cDNA; 1679 BP.		
AC	ABL39772;		
AC			
DT	10-MAY-2002 (first entry)		
XX	Human NS cDNA sequence SEQ ID NO:82.		
XX	Human; cytostatic; osteopathic; gynaecological; neuroprotective; antiarthritis; antiarthritic; antipsoriatic; ophthalmological; anti-HIV; vasotropic; antiarteriosclerotic; antiinflammatory; dermatological; anorectic; muscular; antinfertility; cardiovascular; anticoagulant; antibrinolytic; hypotension; antiasthmatic; immunomodulator; cardiast; anticonvulsant; antidiabetic; tranquiliser; antidepressant; neuroleptic; gastrointestinal; virucide; antiulcer; cerebroprotective; nootropic; contraceptive; vaccine; gene therapy; cancer; osteoporosis; dystonia; endometriosis; degenerative disease; multiple sclerosis; psoriasis; rheumatoid arthritis; cataract; retinosis; atherosclerosis; glaucoma; inflammation; skin disorder; obesity; muscular dystrophy; AIDS; infertility; cardiovascular disease; coagulation disease; hypertension; ischaemia; asthma; immune disease; epilepsy; angina; neurodegeneration; diabetes; anxiety; depression; schizophrenia; viral disease; stroke; gastric ulcer; Alzheimer's disease; gene; ss.		
XX	Homo sapiens.		
OS	WO200206315-A2.		
XX	24-JAN-2002.		
XX	17-JUL-2001; 2001WO-IL000653.		
XX	18-JUL-2000; 2000IL-00137345.		
PR	15-DEC-2000; 2000IL-00140354.		
XX	(COMP-) COMPUGEN LTD.		
PA	Mintz L, Freilich S, Bernstein J;		
XX	WPI; 2002-155037/20.		
DR	P-PSDB; AB06118.		
XX	One hundred and twenty eight novel nucleic acid sequences, useful for treating and diagnosing e.g. cancer, asthma and AIDS.		

XX	Claim 1; Page 123; 290pp; English.	
PS	ABL39691 to ABL39818 represent novel human nucleic acid sequences	
XX	encoding the proteins given in ABB06037 to ABB06164. The novel sequences	
CC	(NS) can have cytostatic, osteopathic, gynaecological, neuroprotective,	
CC	antirheumatic, antithrombotic, antipsoriatic, ophthalmological, virucide,	
CC	vasotropic, antiarteriosclerotic, antiinflammatory, dermatological,	
CC	anorectic, muscular, anti-HIV, antiinfertility, cardiovascular,	
CC	anticoagulant, antifibrinolytic, hypotension, antiasthmatic, antiulcer,	
CC	immunomodulator, anticonvulsant, antidiabetic, tranquillisier, cardiulcer,	
CC	anti-depressant, gastrointestinal, antiepileptic,	
CC	neurotic and contraceptive activities. The NS can be used in vaccines,	
CC	gene therapy and antisense therapy. Nucleic acids, expression vectors and	
CC	antibodies from the present invention can be used for treating and	
CC	diagnosing e.g. cancer, osteoporosis, endometriosis, degenerative	
CC	diseases, dystonia, multiple sclerosis, rheumatoid arthritis, psoriasis,	
CC	cataracts, restenosis, atherosclerosis, inflammation, skin disorders,	
CC	glaucoma, obesity, muscular dystrophy, AIDS, infertility, cardiovascular	
CC	disease, coagulation disease, ischaemia, hypertension, asthma, immune	
CC	disease, epilepsy, angina, neurodegeneration, diabetes, anxiety,	
CC	depression, schizophrenia, viral disease, gastric ulcers, stroke,	
CC	Alzheimer's disease and as a contraceptive	
XX	Sequence 1679 BP; 461 A; 443 C; 378 G; 393 T; 0 U; 4 Other;	
SQ	Query Match 61.2%; Score 1475; DB 6; Length 1679;	
	Best Local Similarity 99.8%; Pred. No. 0;	
	Matches 1675; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
QY	694 GCTATGAGAAAGATGCTTGAATGGCAGCAATAAATAGACACACACCCAGCCACC 753	
DB	1 GCTATGAGAAAGATGCTTGAATGGCAGCAATAAATAGACACACACCCAGCCACC 60	
QY	754 ACTCTGTTGTGATGATGAGGATACCTAAGATCTAGAGTCAATGTTTAAATGTC 813	
DB	61 ACCTGTTGTGATGATGAGGATACCTAAGATCTAGAGTCAATGTTTAAATGTC 120	
QY	814 AAGGGCGTTTCAGATACCTTTTGTATCCATATCTTGATATAACATTTGGAGATAAAGGCTGCT 873	
DB	121 AAGGGCGTTTCAGATACCTTTTGTATCCATATCTTGATATAACATTTGGAGATAAAGGCTGCT 180	
QY	874 CAGAGTGTCAAGGATGAGGAGTGTGAGCGGACAGCTTGTAGGAGAAAC 933	
DB	181 CAGAGTGTCAAGGATGAGGAGTGTGAGCGGACAGCTTGTAGGAGAAAC 240	
QY	934 TCGTACAGTGCAGCAAGTGTAAAGATGTTCCAGCTTCAAGAGAGTTCACTATCCAT 993	
DB	241 TCGTACAGTGCAGCAAGTGTAAAGATGTTCCAGCTTCAAGAGAGTTCACTATCCAT 300	
QY	994 AGATCCTCTAATGTTCTTACACTTTTCTGAAAAGTGTGAAATTTTACCGTGGAAAA 1053	
DB	301 AGATCCTCTAATGTTCTTACACTTTTCTGAAAAGTGTGAAATTTTACCGTGGAAAA 360	
QY	1054 ATTGTAGAGTGTGAAATACCTGAGTATCTGATATTCGSCCATATATGTTCTCAACCC 1113	
DB	361 ATTGTAGAGTGTGAAATACCTGAGTATCTGATATTCGSCCATATATGTTCTCAACCC 420	
QY	1114 AACGAGAGCAATTTGTCTACGCTTGTATGAGTGTGTTCCACACTGGTTTAAATGTC 1173	
DB	421 AACGAGAGCAATTTGTCTACGCTTGTATGAGTGTGTTCCACACTGGTTTAAATGTC 480	
QY	1174 CATGCTGCCATTTACTTCTGCTACATAAAGAGTACCAATGGCTCTGTTATCAATGAAT 1233	
DB	481 CATGCTGCCATTTACTTCTGCTACATAAAGAGTACCAATGGCTCTGTTATCAATGAAT 540	
QY	1234 GACTCCATTTGATCTTACAGTATATAGATCGGTACTCAGCAACAAGCTTATGTGTC 1293	
DB	541 GACTCCATTTGATCTTACAGTATATAGATCGGTACTCAGCAACAAGCTTATGTGTC 600	
QY	1294 TTTTATATCAGGTCCTCCATGATGAAAAATGGAGGTGAATTTACTATCCCAACCATAGC 1353	
DB	601 TTTTATATCAGGTCCTCCATGATGAAAAATGGAGGTGAATTTACTATCCCAACCATAGC 660	

RESULT 8
AAK79798
ID AAK79798 standard; DNA; 16682 BP.

QY	1354 CCGGCGAGTCTCTCCCGCCCGCTCATCAGTCAGCGGTTGTCAACAAACAGGCT 1413	
DB	661 CCGGCGAGTCTCTCTCCCGCCCGCTCATCAGTCAGCGGTTGTCAACAAACAGGCT 720	
QY	1414 GCSCCAGGCTTTATCGGACACACAGCTTCCCTCTCATATGATAAAGAAATCAACCTCACTTA 1473	
DB	721 GCSCCAGGCTTTATCGGACACACAGCTTCCCTCTCATATGATAAAGAAATCAACCTCACTTA 780	
QY	1474 AATGGAGTCTGGACCATTTGAAAGACAGCCCAAGCAGTTCCATGTCGAGTCTCAAGGGAAT 1533	
DB	781 AATGGAGTCTGGACCATTTGAAAGACAGCCCAAGCAGTTCCATGTCGAGTCTCAAGGGAAT 840	
QY	1534 TCCAGTGTCAACAGGCTAGTCTGTTTAAATGTTTCACTTCTGTCACAAATCTGTCAGTT 1593	
DB	841 TCCAGTGTCAACAGGCTAGTCTGTTTAAATGTTTCACTTCTGTCACAAATCTGTCAGTT 900	
QY	1594 AATAGTCTCTCAGTATCCAGAACATCTTAAGAAACAAAAATTAACAATCAGTATTCAC 1653	
DB	901 AATAGTCTCTCAGTATCCAGAACATCTTAAGAAACAAAAATTAACAATCAGTATTCAC 960	
QY	1654 AACAGTGTGCTGTTCCGAGTGTCACTCAACCTCACTTCACTAGTAAATTTCTTTGAG 1713	
DB	961 AACAGTGTGCTGTTCCGAGTGTCACTCAACCTCACTTCACTAGTAAATTTCTTTGAG 1020	
QY	1714 AACCTTACCAAGCCGTTCCCTCTTCTACCAATACCAATTTCTGCAGTACAGTCTACCTCG 1773	
DB	1021 AACCTTACCAAGCCGTTCCCTCTTCTACCAATTTCTGCAGTACAGTCTACCTCG 1080	
QY	1774 AACGATCTACGATCTCAGTCTTCTAGTAAAGTAAACAAACCGATCCCGCGAGTGAATCC 1833	
DB	1081 AACGATCTACGATCTCAGTCTTCTAGTAAAGTAAACAAACCGATCCCGCGAGTGAATCC 1140	
QY	1834 TGCTCCAGCCGCTGATGAATGGCAATCCAGCTCAACTCCAGCTGCTGTCGCCCTAT 1893	
DB	1141 TGCTCCAGCCGCTGATGAATGGCAATCCAGCTCAACTCCAGCTGCTGTCGCCCTAT 1200	
QY	1894 GGCGCCGAGTCTCTGAGGACTCTGACGAGAGTCAAGGGGCTGGGCAAGAGAGATGGG 1953	
DB	1201 GGCGCCGAGTCTCTGAGGACTCTGACGAGAGTCAAGGGGCTGGGCAAGAGAGATGGG 1260	
QY	1954 ATTGTTACGATTTGAGTCTCCACTCTCCCGGCAAGATGCCAGAGTACGAGAGGCACT 2013	
DB	1261 ATTGTTACGATTTGAGTCTCCACTCTCCCGGCAAGATGCCAGAGTACGAGAGGCACT 1320	
QY	2014 CCGCACGAGTTCAGAACCCATGACCTTAAACGCTCTAATAGTGCAGACAGACAGT 2073	
DB	1321 CCGCACGAGTTCAGAACCCATGACCTTAAACGCTCTAATAGTGCAGACAGACAGT 1380	
QY	2074 GACCCGAAAGAAAAACGGCTAGCGCTAGCGCTGATGGTGCAGCTGCCAAGGCCAGCGCTCGCTG 2133	
DB	1381 GACCCGAAAGAAAAACGGCTAGCGCTAGCGCTGATGGTGCAGCTGCCAAGGCCAGCGCTCGCTG 1440	
QY	2134 CACTCAGAAAAATCTTTTGTGTAAGGCAACCGTCTCTCTGAAAGTTGATGCTGCTCTCT 2193	
DB	1441 CACTCAGAAAAATCTTTTGTGTAAGGCAACCGTCTCTCTGAAAGTTGATGCTGCTCTCT 1500	
QY	2194 TTGCTGTCTCTCCAGAGACAAATCTTAGAGACTTTAGAGCTTTAGCAACAACTGAAA 2253	
DB	1501 TTGCTGTCTCTCCAGAGACAAATCTTAGAGACTTTAGAGCTTTAGCAACAACTGAAA 1560	
QY	2254 GGCTCGAGCGATGAATGAGTGCACCTGGAGCAGAGAGGGGCTCTCCGAGAGACCGGAC 2313	
DB	1561 GGCTCGAGCGATGAATGAGTGCACCTGGAGCAGAGAGGGGCTCTCCGAGAGACCGGAC 1620	
QY	2314 GCGAGCTCTAGCTTGGAGCCCGCCCGGAGTCCCTTGGAGGAGCCGAGATGCGGCGC 2372	
DB	1621 GCGAGCTCTAGCTTGGAGCCCGCCCGGAGTCCCTTGGAGGAGCCGAGATGCGGCGC 1679	

XX	AAK79798;	PR	14-SEP-2000;	2000US-02324000;
AC		PR	14-SEP-2000;	2000US-0232401P
XX	07-NOV-2001 (first entry)	PR	14-SEP-2000;	2000US-0233063P
DT		PR	14-SEP-2000;	2000US-0233064P
XX	Human immune/haematopoietic antigen genomic sequence	PR	14-SEP-2000;	2000US-0233065P
DE		PR	21-SEP-2000;	2000US-0234223P
XX		PR	21-SEP-2000;	2000US-0234224P
XX	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;	PR	25-SEP-2000;	2000US-0234997P
KW	cytostatic; gene therapy; vaccine; metastasis; ds.	PR	25-SEP-2000;	2000US-0234998P
KW		PR	26-SEP-2000;	2000US-0234844P
OS	Homo sapiens.	PR	27-SEP-2000;	2000US-0235834P
XX		PR	27-SEP-2000;	2000US-0235836P
XX	WO2000157182-A2.	PR	29-SEP-2000;	2000US-0236327P
PN		PR	29-SEP-2000;	2000US-0236367P
PD	09-AUG-2001.	PR	29-SEP-2000;	2000US-0236368P
XX		PR	29-SEP-2000;	2000US-0236369P
XX		PR	29-SEP-2000;	2000US-0236370P
XX	17-JAN-2001; 2001WO-US0001354.	PR	02-OCT-2000;	2000US-0236802P
31	JAN-2000; 2000US-0179065P.	PR	02-OCT-2000;	2000US-0237037P
PR	04-FEB-2000; 2000US-0180628P.	PR	02-OCT-2000;	2000US-0237038P
PR	24-FEB-2000; 2000US-0184664P.	PR	02-OCT-2000;	2000US-0237039P
PR	02-MAR-2000; 2000US-0186350P.	PR	13-OCT-2000;	2000US-0239935P
PR	16-MAR-2000; 2000US-0189874P.	PR	13-OCT-2000;	2000US-0239937P
PR	17-MAR-2000; 2000US-0190076P.	PR	20-OCT-2000;	2000US-0240960P
PR	18-APR-2000; 2000US-0198123P.	PR	20-OCT-2000;	2000US-0241221P
PR	19-MAY-2000; 2000US-0205515P.	PR	20-OCT-2000;	2000US-0241785P
PR	07-JUN-2000; 2000US-0209467P.	PR	20-OCT-2000;	2000US-0241786P
PR	28-JUN-2000; 2000US-0214886P.	PR	20-OCT-2000;	2000US-0241787P
PR	30-JUN-2000; 2000US-0215135P.	PR	20-OCT-2000;	2000US-0241808P
PR	07-JUL-2000; 2000US-0216647P.	PR	20-OCT-2000;	2000US-0241809P
PR	07-JUL-2000; 2000US-0216880P.	PR	20-OCT-2000;	2000US-0241826P
PR	11-JUL-2000; 2000US-0217487P.	PR	01-NOV-2000;	2000US-0244617P
PR	11-JUL-2000; 2000US-0217496P.	PR	08-NOV-2000;	2000US-0246474P
PR	14-JUL-2000; 2000US-0218290P.	PR	08-NOV-2000;	2000US-0246475P
PR	26-JUL-2000; 2000US-0220963P.	PR	08-NOV-2000;	2000US-0246476P
PR	26-JUL-2000; 2000US-0220964P.	PR	08-NOV-2000;	2000US-0246477P
PR	14-AUG-2000; 2000US-0224518P.	PR	08-NOV-2000;	2000US-0246478P
PR	14-AUG-2000; 2000US-0224519P.	PR	08-NOV-2000;	2000US-0246523P
PR	14-AUG-2000; 2000US-0225211P.	PR	08-NOV-2000;	2000US-0246525P
PR	14-AUG-2000; 2000US-0225214P.	PR	08-NOV-2000;	2000US-0246526P
PR	14-AUG-2000; 2000US-0225266P.	PR	08-NOV-2000;	2000US-0246527P
PR	14-AUG-2000; 2000US-0225267P.	PR	08-NOV-2000;	2000US-0246528P
PR	14-AUG-2000; 2000US-0225268P.	PR	08-NOV-2000;	2000US-0246532P
PR	14-AUG-2000; 2000US-0225270P.	PR	08-NOV-2000;	2000US-0246609P
PR	14-AUG-2000; 2000US-0225447P.	PR	08-NOV-2000;	2000US-0246610P
PR	14-AUG-2000; 2000US-0225447P.	PR	08-NOV-2000;	2000US-0246611P
PR	14-AUG-2000; 2000US-0225757P.	PR	17-NOV-2000;	2000US-0249207P
PR	14-AUG-2000; 2000US-0225758P.	PR	17-NOV-2000;	2000US-0249208P
PR	18-AUG-2000; 2000US-0225759P.	PR	17-NOV-2000;	2000US-0249210P
PR	22-AUG-2000; 2000US-0226279P.	PR	17-NOV-2000;	2000US-0249211P
PR	22-AUG-2000; 2000US-0226681P.	PR	17-NOV-2000;	2000US-0249212P
PR	22-AUG-2000; 2000US-0226686P.	PR	17-NOV-2000;	2000US-0249213P
PR	23-AUG-2000; 2000US-0227182P.	PR	17-NOV-2000;	2000US-0249214P
PR	30-AUG-2000; 2000US-0228924P.	PR	17-NOV-2000;	2000US-0249215P
PR	01-SEP-2000; 2000US-0229287P.	PR	17-NOV-2000;	2000US-0249216P
PR	01-SEP-2000; 2000US-0229343P.	PR	17-NOV-2000;	2000US-0249

PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Disclosure; SEQ ID NO 34610; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
XX Sequence 16682 BP; 4217 A; 3754 C; 4167 G; 4538 T; 0 U; 6 Other;
SQ
Query Match 29.9%; Score 722; DB 4; Length 16682;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 722; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1457 AGAATCCACCTCACTTAAATGGGACTGGACCAATTAAGAGACACGCCAAGCAGTTCATGT 1516
Db 9451 AGAATCCACCTCACTTAAATGGGACTGGACCAATTAAGAGACACGCCAAGCAGTTCATGT 9510
Qy 1517 CGAGTCTCTAACGGGAATCCAGTGTCAACAGGGCTAGTCCCTGTTAAATGCTTCAGCTTCTG 1576
Db 9511 CGAGTCTCTAACGGGAATCCAGTGTCAACAGGGCTAGTCCCTGTTAAATGCTTCAGCTTCTG 9570
Qy 1577 TCCAAAATCTGCTCAGTTAATAGGTGCTCAGTGATCCCAAGACATCCTTAAGAAACAAAAA 1636
Db 9571 TCCAAAATCTGCTCAGTTAATAGGTGCTCAGTGATCCCAAGACATCCTTAAGAAACAAAAA 9630
Qy 1637 TTCAATCAGTATTCAACAAAGTTGCTCTGTCGCGCAGTGTCAAGTCAACCTAACCTTC 1696
Db 9631 TTCAATCAGTATTCAACAAAGTTGCTCTGTCGCGCAGTGTCAAGTCAACCTAACCTTC 9690
Qy 1697 ATAGTAATCTTTGAGAACCTTACCAAGCCGTTCCCTCTTACCAATTAACCAATTCG 1756
Db 9691 ATAGTAATCTTTGAGAACCTTACCAAGCCGTTCCCTCTTACCAATTAACCAATTCG 9750
Qy 1757 CAGTACAGTCTACCTCGAAGCGATCTACGATGTCAAGTGTCTAGTAAAGTAACAAACCGA 1816
Db 9751 CAGTACAGTCTACCTCGAAGCGATCTACGATGTCAAGTGTCTAGTAAAGTAACAAACCGA 9810
Qy 1817 TCCCCCGAGTGAATCTCTGCTCCGAGCCGTCATGAATGGCAAAATCCAAAGCTGAATCCA 1876
Db 9811 TCCCCCGAGTGAATCTCTGCTCCGAGCCGTCATGAATGGCAAAATCCAAAGCTGAATCCA 9870
Qy 1877 GCGTGTCTGTCCTATGCGCGCGAGTCTCTGAGGACTCTGACGAGGAGTCAAAAGGGGC 1936

Db 9871 GCGTGTCTGTCCTATGCGCGCGAGTCTCTGAGGACTCTGACGAGGAGTCAAAAGGGGC 9930
Qy 1937 TGGGCAAGGAGGAGTGGATTGGTACGATTGTGAGTCTCCCACTCTCCCGGCCAAGATGCCG 1996
Db 9931 TGGGCAAGGAGGAGTGGATTGGTACGATTGTGAGTCTCCCACTCTCCCGGCCAAGATGCCG 9990
Qy 1997 AAGATCAGGAGGCACTCCGACGAGCTTCAAGAACCCATGACCTTAACCGGTGCTTAATA 2056
Db 9991 AAGATCAGGAGGCACTCCGACGAGCTTCAAGAACCCATGACCTTAACCGGTGCTTAATA 10050
Qy 2057 GTGCAGACAGCGACAGTACCCGAAAAAAGAAACCGCCCTAGCGCTGATGTCGCGAGTGC 2116
Db 10051 GTGCAGACAGCGACAGTACCCGAAAAAAGAAACCGCCCTAGCGCTGATGTCGCGAGTGC 10110
Qy 2117 AAGCCAGCTGCGCTGCACTCAGAAATCCCTTTGCTAAGGCAAAACGGTCTTCTCTGGAA 2176
Db 10111 AAGCCAGCTGCGCTGCACTCAGAAATCCCTTTGCTAAGGCAAAACGGTCTTCTCTGGAA 10170
Qy 2177 AG 2178
Db 10171 AG 10172
RESULT 9
AAK70010
ID AAK70010 standard; DNA; 16682 BP.
AC AAK70010;
XX
DT 06-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24822.
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.

Db 9451 AGAATCCACCTCACCTTAATGGAGCTGGACCATTTGAAGACACGCCAAGCAGTTCCATCT 9510
QY 1517 CGAGTCTTAACGGGAATTCAGTGTCAACAGGCTAGTCTGTAAATGTTTCAAGTCTTG 1576
Db 9511 CGAGTCTTAACGGGAATTCAGTGTCAACAGGCTAGTCTGTAAATGTTTCAAGTCTTG 9570
QY 1577 TCCAAACTGGTCAGTAAATAGTCTCTCAGTATCCAGAACATCTTAAGAACAAAAA 1636
Db 9571 TCCAAACTGGTCAGTAAATAGTCTCTCAGTATCCAGAACATCTTAAGAACAAAAA 9630
QY 1637 TTACAATCAGTATTCACAAAGTTGCTCTTCGCGAGTGTCAAGTCTCAACCTTAACCTTC 1696
Db 9631 TTACAATCAGTATTCACAAAGTTGCTCTTCGCGAGTGTCAAGTCTCAACCTTAACCTTC 9690
QY 1697 ATAGTAATCTTTGGAGAACCTTACCAAGCCCTTCCCTTTTACCATTAACCAATCTTG 1756
Db 9691 ATAGTAATCTTTGGAGAACCTTACCAAGCCCTTCCCTTTTACCATTAACCAATCTTG 9750
QY 1757 CAGTACAGTCTACTCGAAGCGATCTACGATGTCTAGTGTCTTCTAGTAAAGTAAACCCGA 1816
Db 9751 CAGTACAGTCTACTCGAAGCGATCTACGATGTCTAGTGTCTTCTAGTAAAGTAAACCCGA 9810
QY 1817 TCCCGCGCAGTGAATCTCTGCTCCAGCCCTGATGAATGGCAATCCAACTGAACCTCA 1876
Db 9811 TCCCGCGCAGTGAATCTCTGCTCCAGCCCTGATGAATGGCAATCCAACTGAACCTCA 9870
QY 1877 GCGTGTGTCCTTATGGCCGAGTCTCTGAGGACTCTGACGAGGATCAAGGGGC 1936
Db 9871 GCGTGTGTCCTTATGGCCGAGTCTCTGAGGACTCTGACGAGGATCAAGGGGC 9930
QY 1937 TGGCAGAGGATGGATGGTACGATTTGTAGCTCCCTCCCGCCCAAGATCCG 1996
Db 9931 TGGCAGAGGATGGATGGTACGATTTGTAGCTCCCTCCCGCCCAAGATCCG 9990
QY 1997 AAGATGAGGAGGACCTCCGACAGCTTCAAGAACCCATGACCTTAAACGGTCTAATA 2056
Db 9991 AAGATGAGGAGGACCTCCGACAGCTTCAAGAACCCATGACCTTAAACGGTCTAATA 10050
QY 2057 GTGCAGACAGCAGTACCCGAAAGAAACCGGCTAGCGCTGATGGTCCAGCTGCC 2116
Db 10051 GTGCAGACAGCAGTACCCGAAAGAAACCGGCTAGCGCTGATGGTCCAGCTGCC 10110
QY 2117 AAGCCAGCTGCTGCTCAGTCAAGAAATCCCTTTGCTAAGGCAAAACGGTCTTCTCGAA 2176
Db 10111 AAGCCAGCTGCTGCTCAGTCAAGAAATCCCTTTGCTAAGGCAAAACGGTCTTCTCGAA 10170
QY 2177 AG 2178
Db 10171 AG 10172

RESULT 10
AAK79799
ID AAK79799 standard; DNA; 21436 BP.
XX
AC AAK79799;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34611.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cystostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR

PR 04-FEB-2000; 2000US-0180528P.
PR 24-FEB-2000; 2000US-0184564P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189374P.
PR 17-MAR-2000; 2000US-0190769P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205315P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225467P.
PR 14-AUG-2000; 2000US-0225468P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 14-AUG-2000; 2000US-0226279P.
PR 18-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228244P.
PR 01-SEP-2000; 2000US-0225287P.
PR 01-SEP-2000; 2000US-0225343P.
PR 01-SEP-2000; 2000US-0225344P.
PR 01-SEP-2000; 2000US-0225345P.
PR 05-SEP-2000; 2000US-0225509P.
PR 05-SEP-2000; 2000US-0225513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0235802P.
PR 02-OCT-2000; 2000US-0237037P.

PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249266P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251983P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-483426/52.

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.

PS Disclosure; SEQ ID NO 34611; 3071pp + Sequence Listing; English.

XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 21436 BP; 5267 A; 4950 C; 5418 G; 5801 T; 0 U; 0 Other;

Query Match 29.9%; Score 722; DB 4; Length 21436;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 722; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1457 AGAATCCACCTCACTTAATGGACCTGGACCATTTGAAGACACGCCAAGCAGTTCCATGT 1516
DB 9454 AGAATCCACCTCACTTAATGGACCTGGACCATTTGAAGACACGCCAAGCAGTTCCATGT 9513
QY 1517 CGAGTCCTTAACGGGAATTCAGTGTCAACAGGGCTAGTCTGTATGCTTCAGCTTCTG 1576
DB 9514 CGAGTCCTTAACGGGAATTCAGTGTCAACAGGGCTAGTCTGTATGCTTCAGCTTCTG 9573
QY 1577 TCCAAAACCTGGTCAGTTAATAGGTCTCTAGTGTATCCCAAGCATCTTAAGAACCAAAAA 1636
DB 9574 TCCAAAACCTGGTCAGTTAATAGGTCTCTAGTGTATCCCAAGCATCTTAAGAACCAAAAA 9633
QY 1637 TTACAATCAGTATTCAACAAGTTGCCCTGTTCCGACGTGTGAGTCTCAACCTTAACTTC 1696
DB 9634 TTACAATCAGTATTCAACAAGTTGCCCTGTTCCGACGTGTGAGTCTCAACCTTAACTTC 9693
QY 1697 ATAGTAATTTCTTGAGAACCTTACCAAGCCCGTTCCTCTTCTTACCATTAACAATCTG 1756
DB 9694 ATAGTAATTTCTTGAGAACCTTACCAAGCCCGTTCCTCTTCTTACCATTAACAATCTG 9753
QY 1757 CAGTACAGTCTACCTCGAAACGCATCTACGATGTCTAGTAAAGTAAACAAACCGA 1816
DB 9754 CAGTACAGTCTACCTCGAAACGCATCTACGATGTCTAGTAAAGTAAACAAACCGA 9813
QY 1817 TCCCCCGCAGTGAATCTCTGCTCCAGCCCGTGAATGGCAAAATCCAAGCTGAATCCA 1876
DB 9814 TCCCCCGCAGTGAATCTCTGCTCCAGCCCGTGAATGGCAAAATCCAAGCTGAATCCA 9873
QY 1877 GCGTGTGCTGCTGCTATGGCCGAGTCTCTGAGGACTCTGACGAGAGTCAAGGGGC 1936
DB 9874 GCGTGTGCTGCTGCTATGGCCGAGTCTCTGAGGACTCTGACGAGAGTCAAGGGGC 9933
QY 1937 TGGGCAAGGAGAATGGGAATGGTACGATTGTGAGCTCCCACTCTCCCGGCCAAGATGCG 1996
DB 9934 TGGGCAAGGAGAATGGGAATGGTACGATTGTGAGCTCCCACTCTCCCGGCCAAGATGCG 9993
QY 1997 AAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCTTAAACGGTCTAATA 2056
DB 9994 AAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCTTAAACGGTCTAATA 10053
QY 2057 GTGCAGACAGCGACAGTGAACCGGAAGAAACGGCTTAGCCCTGATGTGTCAGTGTGCC 2116
DB 10054 GTGCAGACAGCGACAGTGAACCGGAAGAAACGGCTTAGCCCTGATGTGTCAGTGTGCC 10113
QY 2117 AAGGCCAGCTGCCCTGCACCTCAGAAAAATCCCTTTGCTTAAGGCAAAACGGTCTTCTCGAA 2176
DB 10114 AAGGCCAGCTGCCCTGCACCTCAGAAAAATCCCTTTGCTTAAGGCAAAACGGTCTTCTCGAA 10173

Wed Aug 18 13:52:58 2004

us-10-049-745-31.oli30.rng

```
QY      2177 AG 2178
Db      10174 AG 10175

RESULT 11
AAK70011
ID AAK70011 standard; DNA; 21436 BP.
AC AAK70011;
XX
XX
XX
DT 06-NOV-2001 (first entry)
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24823.
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0217496P.
PR 26-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229387P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0229437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232380P.
PR 08-SEP-2000; 2000US-0232381P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
```


Novel polypeptides and polynucleotides useful for diagnosing, preventing, treating neural, immune system, muscular, reproductive, pulmonary, cardiovascular, renal, proliferative disorders and cancerous diseases.

Claim 4; SEQ ID NO 839; 1180pp; English.

The present invention relates to the isolation of novel human enzyme polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences encoding them. The enzyme polypeptides of the invention may comprise the functional classes of oxidoreductases, transferases, hydrolases, lyases, isomerases or ligases. The sequences of the invention are useful in the diagnosis, treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS), autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis), blood-related disorders (e.g. haemophilia), reproductive disorders (e.g. infertility) and infectious disorders (e.g. influenza). The polynucleotides of the invention can also be used in gene therapy. AAU40785-AAU41684 represent cDNA sequences encoding for the novel human enzyme polypeptides of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 577 BP; 138 A; 138 C; 147 G; 151 T; 0 U; 3 Other;

Query Match 17.6%; Score 424; DB 4; Length 577;
Best Local Similarity 100.0%; Pred. No. 3 4e-197;
Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	35	GGGATGGAGGAGCGCGCGGTCAGAGTTGAAACAAATGACCATAGTTGACAAAGCTT	94
DB	46	GGGATGGAGGAGCGCGCGGTCAGAGTTGAAACAAATGACCATAGTTGACAAAGCTT	105
QY	95	CTGAATCTTCAGACCCATCAGCCTATCAGATCAGCCTGCGAGCTCCGAGGAGTCTCAC	154
DB	106	CTGAATCTTCAGACCCATCAGCCTATCAGATCAGCCTGCGAGCTCCGAGGAGTCTCAC	165
QY	155	CTGGAGACATGGATGAGTTCTGCCAGCTGGGCTGCTGCTTCTTCAATGATGATGT	214
DB	166	CTGGAGACATGGATGAGTTCTGCCAGCTGGGCTGCTGCTTCTTCAATGATGATGT	225
QY	215	CAAAATCACACATTTCTTTAGGACAGTACCTGGTGTCTAGTTTATTCGGATTCATCTG	274
DB	226	CAAAATCACACATTTCTTTAGGACAGTACCTGGTGTCTAGTTTATTCGGATTCATCTG	285
QY	275	TACCTGATAAATCAAAACCATCACCAAAAGGATCAAGCCCTAGCTGATGGCATCGCTC	334
DB	286	TACCTGATAAATCAAAACCATCACCAAAAGGATCAAGCCCTAGCTGATGGCATCGCTC	345
QY	335	CTCCACAGAAAGTTCTTTTCCCATCTGAGAAGATTGTCTTAAGTGGCAACAACTCATA	394
DB	346	CTCCACAGAAAGTTCTTTTCCCATCTGAGAAGATTGTCTTAAGTGGCAACAACTCATA	405
QY	395	CAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTTCCCAATGAGCACTGCAGT	454
DB	406	GAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTTCCCAATGAGCACTGCAGT	465
QY	455	GTTT 458	
DB	466	GTTT 469	

RESULT 14
AAI92171
ID AAI92171 standard; cDNA; 411 BP.
XX
AC AAI92171;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 12231.

29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-465566/50.
P-PSDB; AAU23743.
XX
XX
XX
XX
XX
XX

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX Homo sapiens.
OS
XX WO200164835-A2.
PN
XX 07-SEP-2001.
PD
XX
PF 26-FEB-2001; 2001WO-US004927.
XX
PR 28-FEB-2000; 2000US-00515126.
PR 18-MAY-2000; 2000US-00577409.
XX
XX (HYSE-) HYSEQ INC.
PA
PI Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
DR P-PSDB; AA012240.
DR
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
PT
XX Claim 1; SEQ ID NO 12231; 1399pp + Sequence Listing; English.
PS
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 411 BP; 108 A; 99 C; 99 G; 105 T; 0 U; 0 Other;

Query Match
Best Local Similarity 17.0%; Score 411; DB 4; Length 411;
Matches 411; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 GGATGGAGCGAGCGCGGTCAGAGTTGAAACATGACCAATGACCAAGCTTC 95
DB 1 GGATGGAGCGAGCGCGGTCAGAGTTGAAACATGACCAATGACCAAGCTTC 60
QY 96 TGAATCTTCAGACCCATCAGCTATCAGAAATCAGCTGCGAGCTCCGAGGAGTCTCACC 155
DB 61 TGAATCTTCAGACCCATCAGCTATCAGAAATCAGCTGCGAGCTCCGAGGAGTCTCACC 120
QY 156 TGGAGAGATGATCAGGTTCTGCGAGCTGCGGCTGCTGCTTCATGAAATGATGTC 215
DB 121 TGGAGAGATGATCAGGTTCTGCGAGCTGCGGCTGCTGCTTCATGAAATGATGTC 180
QY 216 AAATCACACACTTCTTAGACACAGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 275
DB 181 AAATCACACACTTCTTAGACACAGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
QY 276 ACTGTAAATCAAAACCATCACCACAAAGGATCAAGCCCTAGGTGATGCGATCGCTCC 335
DB 241 ACTGTAAATCAAAACCATCACCACAAAGGATCAAGCCCTAGGTGATGCGATCGCTCC 300
QY 336 TCCACAGAAAGTTCTTTCCATCTGAGAGATTTGCTTAAATGGCAACAACTCATAG 395
DB 301 TCCACAGAAAGTTCTTTCCATCTGAGAGATTTGCTTAAATGGCAACAACTCATAG 360

QY 396 AGTTGAGCTGGCTCCAGAAATTTGGCAATACCTGTTTGGCAATGAGC 446
DB 361 AGTTGAGCTGGCTCCAGAAATTTGGCAATACCTGTTTGGCAATGAGC 411
RESULT 15
ABK30453
ID ABK30453 standard; cDNA; 722 BP.
XX
AC ABK30453;
XX
DT 23-APR-2002 (first entry)
XX
DE Human G-protein-coupled protease #223.
XX
KW Human; ss; gene; G-protein-coupled protease; gene therapy; transgenic;
KW protease mediated disorder; proliferative disorder;
KW differentiative disorder; developmental disorder;
KW haematopoietic disorder.
XX
OS Homo sapiens.
XX
PN US6331427-B1.
XX
PD 18-DEC-2001.
XX
PF 26-MAR-1999; 99US-00280116.
XX
PR 26-MAR-1999; 99US-00280116.
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Robison KE;
XX
DR WPI; 2002-129545/17.
XX
XX New polynucleotides encoding protease homologs of the G-protein-coupled
PT protease family, useful in identifying agonists and antagonists for
PT diagnosis and treatment of protease mediated disorders.
XX
PS Disclosure; Col 251-252; 246pp; English.
XX
CC The invention relates to an isolated human protease nucleic acid molecule
CC comprising a nucleotide sequence of 546 base pairs, one of 268 fully
CC defined in the specification. Also disclosed are production of an
CC isolated polypeptide encoded by the nucleic acid, comprising introducing
CC the nucleic acid into a host cell and culturing under conditions to
CC express the protein from the nucleic acid, use of an antibody to detect
CC the encoded protein in a sample and to modulate its in vivo activity,
CC identifying agents that bind to the protein and identification of a
CC polynucleotide agent that modulates the expression of the nucleic acid or
CC its complement (i.e. gene therapy). The nucleic acid can be used to
CC identify an agent that modulates the expression or activity of the
CC nucleic acid, and can be used to isolate the protein. The nucleic acid
CC can be used in diagnostic assays for determining nucleic acid expression
CC as well as activity in the context of a biological sample (e.g., blood,
CC serum, cells, tissue) to determine whether an individual has a disease or
CC disorder, or is at risk of developing a disease or disorder, associated
CC with aberrant expression or activity of the nucleic acid. The nucleic
CC acid can be used to detect mutations in protease genes and gene
CC expression products such as mRNA. The nucleic acid can be used as
CC hybridisation probes to detect naturally-occurring genetic mutations in a
CC protease gene. The nucleic acid can be used in drug screening methods to
CC identify agonists and antagonists that can be used to diagnose and treat
CC such protease mediated disorders e.g., proliferative, differentiative,
CC developmental or haematopoietic disorders. The nucleic acid can be used
CC as probes, primers, in biological assays, to determine patterns of gene
CC expression, to design ribozymes and to construct transgenic animals. The
CC present sequence represents one of the 268 disclosed human G-protein-
CC coupled protease cDNA sequences
XX
SQ Sequence 722 BP; 189 A; 186 C; 145 G; 196 T; 0 U; 6 Other;

Query Match 14.8%; Score 357; DB 6; Length 722;
Best Local Similarity 100.0%; Pred. No. 2.5e-164; Indels 0; Gaps 0;
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1061 AGGATGTAATACCTGAGTATCTTGATATTCGGCCATATATGCTCAACCCAAACGGAG 1120
107 AGGATGTAATACCTGAGTATCTTGATATTCGGCCATATATGCTCAACCCAAACGGAG 166
1121 AGCCAAATCTCTACGCTCTTGATATGAGTCTGCTCCACACTGGTTTAAATGCCATGCTG 1180
167 AGCCAAATCTCTACGCTCTTGATATGAGTCTGCTCCACACTGGTTTAAATGCCATGCTG 226
1181 GCATTAATCTCTACATATAAAGCTAGCAATGGCTCTGGTATCAAAATGAATGACTCCA 1240
227 GCATTAATCTCTACATATAAAGCTAGCAATGGCTCTGGTATCAAAATGAATGACTCCA 286
1241 TTGTATCTACCAAGTATATAGATCGGTACTCAGCCCAACAAAGCTATGCTCTTTTATA 1300
287 TTGTATCTACCAAGTATATAGATCGGTACTCAGCCCAACAAAGCTATGCTCTTTTATA 346
1301 TCAGGTCCCATGATGTAATAATGAGGTGAATTAATCCACCCATAGCCCGGCC 1360
347 TCAGGTCCCATGATGTAATAATGAGGTGAATTAATCCACCCATAGCCCGGCC 406
1361 AGTCTCTCTCCCGCCCGCTCATGCTCAGCGGTTGTCACCAACAAACAGGCTGCC 1417
407 AGTCTCTCTCCCGCCCGCTCATGCTCAGCGGTTGTCACCAACAAACAGGCTGCC 463

RESULT 16
AAI22307/c
ID AAI22307 standard; DNA; 333 BP.
XX AC AAI22307;
XX DT 12-OCT-2001 (first entry)

Probe #12240 for gene expression analysis in human cervical cell sample.
XX DE Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer; ss.
XX OS Homo sapiens.

XX PN WO200157278-A2.
XX PD 09-AUG-2001.
XX XX 30-JAN-2001; 2001WO-US000670.
XX PF 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX FI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-488901/53.
XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human cervical epithelial cells.
XX PS Claim 25; SEQ ID NO 12240; 487pp; English.

XX XX The present invention relates to human single exon nucleic acid probes
XX CC (SENPs). The present sequence is one such probe. The SENPs are derived
XX CC from human HeLa cells. The SENPs can be used to produce a single exon
XX CC microarray, which can be used for measuring human gene expression in a

CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging of
CC diseases of the cervix, notably cervical cancer. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 333 BP; 57 A; 96 C; 95 G; 85 T; 0 U; 0 Other;

Query Match 13.8%; Score 333; DB 4; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.5e-152; Indels 0; Gaps 0;
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1843 CCCGTGATGAATGGCAATCCAAAGCTGAACTCCAGCGTCTGCTGATGGCGCCGAG 1902
DB 333 CCCGTGATGAATGGCAATCCAAAGCTGAACTCCAGCGTCTGCTGATGGCGCCGAG 274

QY 1903 TCTCTGAGAGCTCTTACGAGAGTCAAAAGGGCTGGCAAGAGAGATGGATTGCTACG 1962
DB 273 TCTCTGAGAGCTCTTACGAGAGTCAAAAGGGCTGGCAAGAGAGATGGATTGCTACG 214

QY 1963 ATTGTGAGCTCCACACTCTCCCGCCAAAGATGCCGAAGATGAGGAGGCCACTCCGACGAG 2022
DB 213 ATTGTGAGCTCCACACTCTCCCGCCAAAGATGCCGAAGATGAGGAGGCCACTCCGACGAG 154

QY 2023 CTTCAAGAAACCCATGACCCCTTAAACCGTGTCTAATAGTGCAGACAGCAGTACCCGAAA 2082
DB 153 CTTCAAGAAACCCATGACCCCTTAAACCGTGTCTAATAGTGCAGACAGCAGTACCCGAAA 94

QY 2083 GAAACGGCGCTAGCGCTGATGTGTCAGTGCAGTGCAGAGCCAGCCCTGCTGCACTCAGAA 2142
DB 93 GAAACGGCGCTAGCGCTGATGTGTCAGTGCAGTGCAGAGCCAGCCCTGCTGCACTCAGAA 34

QY 2143 AATCCCTTTGCTAAGCAACCGTCTTCTCTGGA 2175
DB 33 AATCCCTTTGCTAAGCAACCGTCTTCTCTGGA 1

RESULT 17
ABA67385/c
ID ABA67385 standard; DNA; 333 BP.
XX AC ABA67385;
XX DT 01-FEB-2002 (first entry)
XX XX Human foetal liver single exon nucleic acid probe #15690.
XX DE Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX KW Homo sapiens.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-483447/52.

XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing

PT gene expression in human fetal liver.
 XX
 PS Claim 4; SEQ ID NO 15690; 639pp + Sequence Listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for measuring
 CC human gene expression in a sample derived from human foetal liver. The
 CC single exon nucleic acid probes may be used for predicting, measuring and
 CC displaying gene expression in samples derived from human fetal liver. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences

XX
 SQ Sequence 333 BP; 57 A; 96 C; 95 G; 85 T; 0 U; 0 Other;
 Query Match 13.8%; Score 333; DB 4; Length 333;
 Best Local Similarity 100.0%; Pred. No. 1.5e-152;
 Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1843 CCCGTGATGAATGGCAAAATCCAAAGCTGAACTCCAGCGTGTGCTGCTATGCGCCGAG 1902
 DB 333 CCCGTGATGAATGGCAAAATCCAAAGCTGAACTCCAGCGTGTGCTGCTATGCGCCGAG 274
 OY 1903 TCCTCTGAGGACTCTGACGAGGAGTCAAAGGGCTGGGCAAGAGATGGGATTGGTAGC 1962
 DB 273 TCCTCTGAGGACTCTGACGAGGAGTCAAAGGGCTGGGCAAGAGATGGGATTGGTAGC 214
 OY 1963 ATTGTGAGTCCCACTCTCCCGGCCAAGATGCCAGATGAGAGGCCACTCGCACAGAG 2022
 DB 213 ATTGTGAGTCCCACTCTCTCCCGGCCAAGATGCCAGATGAGAGGCCACTCGCACAGAG 154
 OY 2023 CTTCAAGAACCCATGACCCCTAAACGGTGTCTAATAGTGACAGACGACAGTCCCGGAA 2082
 DB 153 CTTCAAGAACCCATGACCCCTAAACGGTGTCTAATAGTGACAGACGACAGTCCCGGAA 94
 OY 2083 GAAACGGCTTAGCGCTGATGGTCCAGCTGCCAAGGCCAGCCCTGCCCTGCACTCAGAA 2142
 DB 93 GAAACGGCTTAGCGCTGATGGTCCAGCTGCCAAGGCCAGCCCTGCCCTGCACTCAGAA 34
 OY 2143 AATCCCTTTGCTAAGGCAACGGTCTTCTCTGGA 2175
 DB 33 AATCCCTTTGCTAAGGCAACGGTCTTCTCTGGA 1

RESULT 18
 ABA49470/c
 ID ABA49470 standard; DNA; 333 BP.
 XX
 AC ABA49470;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human breast cell single exon nucleic acid probe #8165.
 XX
 KW Human; microarray; single exon probe; gene expression; breast; disease;
 KW cancer; ss.
 XX
 OS Homo sapiens.
 XX
 EN WO200157271-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000662.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-496933/54.
 XX
 PT New spatially-addressable set of single exon nucleic acid probes, useful
 PT for measuring gene expression in sample derived from human breast,
 XX comprises number of single exon nucleic acid probes.
 XX
 PS Claim 4; SEQ ID NO 8165; 327pp + Sequence Listing; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting the
 CC probes with a collection of detectably labelled nucleic acids derived
 CC from mRNA of the human breast, and then measuring the label bound to each
 CC probe of the microarray. The probes are useful for verifying the
 CC expression of regions of genomic DNA predicted to encode proteins. They
 CC are useful for gene discovery, and for determining predisposition and/or
 CC prognosing breast disease. Gene expression analysis is useful for
 CC assessing the toxicity of chemical agents on cells. The microarray of
 CC this invention presents a far greater diversity of probes for measuring
 CC gene expression, with far less bias than expressed sequence tag
 CC microarrays. The method is suitable for rapid production of functional
 CC information from genomic sequence. The present sequence is a single exon
 CC nucleic acid probe of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX
 SQ Sequence 333 BP; 57 A; 96 C; 95 G; 85 T; 0 U; 0 Other;
 Query Match 13.8%; Score 333; DB 4; Length 333;
 Best Local Similarity 100.0%; Pred. No. 1.5e-152;
 Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1843 CCCGTGATGAATGGCAAAATCCAAAGCTGAACTCCAGCGTGTGCTGCTATGCGCCGAG 1902
 DB 333 CCCGTGATGAATGGCAAAATCCAAAGCTGAACTCCAGCGTGTGCTGCTATGCGCCGAG 274
 OY 1903 TCCTCTGAGGACTCTGACGAGGAGTCAAAGGGCTGGGCAAGAGATGGGATTGGTAGC 1962
 DB 273 TCCTCTGAGGACTCTGACGAGGAGTCAAAGGGCTGGGCAAGAGATGGGATTGGTAGC 214
 OY 1963 ATTGTGAGTCCCACTCTCCCGGCCAAGATGCCAGATGAGAGGCCACTCGCACAGAG 2022
 DB 213 ATTGTGAGTCCCACTCTCTCCCGGCCAAGATGCCAGATGAGAGGCCACTCGCACAGAG 154
 OY 2023 CTTCAAGAACCCATGACCCCTAAACGGTGTCTAATAGTGACAGACGACAGTCCCGGAA 2082
 DB 153 CTTCAAGAACCCATGACCCCTAAACGGTGTCTAATAGTGACAGACGACAGTCCCGGAA 94
 OY 2083 GAAACGGCTTAGCGCTGATGGTCCAGCTGCCAAGGCCAGCCCTGCCCTGCACTCAGAA 2142
 DB 93 GAAACGGCTTAGCGCTGATGGTCCAGCTGCCAAGGCCAGCCCTGCCCTGCACTCAGAA 34
 OY 2143 AATCCCTTTGCTAAGGCAACGGTCTTCTCTGGA 2175
 DB 33 AATCCCTTTGCTAAGGCAACGGTCTTCTCTGGA 1

RESULT 19
 AAK41552/c
 ID AAK41552 standard; DNA; 333 BP.
 XX
 AC AAK41552;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human bone marrow expressed single exon probe SEQ ID NO: 16109.
 XX
 KW Human; bone marrow expressed exon; gene expression analysis; probe;

KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX Homo sapiens.
OS WO200157276-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000668.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
XX Example 4; SEQ ID NO 16109; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention
XX
SQ Sequence 333 BP; 57 A; 96 C; 95 G; 85 T; 0 U; 0 Other;
Query Match 13.8%; Score 333; DB 4; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.5e-152;
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1843 CCGTGATGAATGGCAATCCAACTGAACTCCAGCGTGTGGTCCCTATGGCGCCGAG 1902
Db 333 CCGTGATGAATGGCAATCCAACTGAACTCCAGCGTGTGGTCCCTATGGCGCCGAG 274
QY 1903 TCCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGGAGATGGGATTGGTACG 1962
Db 273 TCCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGGAGATGGGATTGGTACG 214
QY 1963 ATTGTGAGCTCCCACTCTCCCGGCAAGATCCGAGATGAGGAGGCCACTCCGACCGAG 2082
Db 213 ATTGTGAGCTCCCACTCTCCCGGCAAGATCCGAGATGAGGAGGCCACTCCGACCGAG 154
QY 2083 CTTCAAGAACCCATGACCCCTAAACGGTCTAATAGTCAGACAGCAGCAGTACCCGAAA 94
QY 2143 AATCCCTTTGCTAAGGCAAAACGGTCTTCTGGA 2175
Db 93 GAAAACGGCTTAGCGCTGATGGTCCAGCTGCCAAGGCCAGCCTGCCCTCAGTACGAA 2142
33 AATCCCTTTGCTAAGGCAAAACGGTCTTCTGGA 1
RESULT 20
AAI07994/c
ID AAI07994 standard; DNA; 333 BP.
XX
AC AAI07994;
XX

DT 09-OCT-2001 (first entry)
XX Probe #7985 used to measure gene expression in human breast sample.
DE
XX
XX Probe; human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
XX WO200157270-A2.
XX 09-AUG-2001.
XX 29-JAN-2001; 2001WO-US000661.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX Novel single exon nucleic acid probe used to measuring gene expression in
PT a human breast.
PT
PT Claim 25; SEQ ID NO 7985; 322pp; English.
PS
XX The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and non-
CC carcinoma tumours. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 333 BP; 57 A; 96 C; 95 G; 85 T; 0 U; 0 Other;
Query Match 13.8%; Score 333; DB 5; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.5e-152;
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1843 CCGTGATGAATGGCAATCCAACTGAACTCCAGCGTGTGGTCCCTATGGCGCCGAG 1902
Db 333 CCGTGATGAATGGCAATCCAACTGAACTCCAGCGTGTGGTCCCTATGGCGCCGAG 274
QY 1903 TCCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGGAGATGGGATTGGTACG 1962
Db 273 TCCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGGAGATGGGATTGGTACG 214
QY 1963 ATTGTGAGCTCCCACTCTCCCGGCAAGATCCGAGATGAGGAGGCCACTCCGACCGAG 2022
Db 213 ATTGTGAGCTCCCACTCTCCCGGCAAGATCCGAGATGAGGAGGCCACTCCGACCGAG 154
QY 2023 CTTCAAGAACCCATGACCCCTAAACGGTCTAATAGTCAGACAGCAGTACCCGAAA 2082
Db 153 CTTCAAGAACCCATGACCCCTAAACGGTCTAATAGTCAGACAGCAGTACCCGAAA 94
QY 2083 GAAAACGGCTTAGCGCTGATGGTCCAGCTGCCAAGGCCAGCCTGCCCTCAGTACGAA 2142
Db 93 GAAAACGGCTTAGCGCTGATGGTCCAGCTGCCAAGGCCAGCCTGCCCTCAGTACGAA 34

XX	03-AUG-2000; 2000US-00632366.	XX	09-AUG-2001.
XX	21-SEP-2000; 2000US-0234687P.	XX	30-JAN-2001; 2001WO-US0000669.
XX	27-SEP-2000; 2000US-0236359P.	XX	04-FEB-2000; 2000US-0180312P.
XX	04-OCT-2000; 2000GB-00024263.	XX	26-MAY-2000; 2000US-0207456P.
XX	(MOLE-) MOLECULAR DYNAMICS INC.	XX	30-JUN-2000; 2000US-00603408.
XX	Penn SG, Hanzel DK, Chen W, Rank DR;	XX	03-AUG-2000; 2000US-00632366.
XX	WPI; 2001-488901/53.	XX	21-SEP-2000; 2000US-0234587P.
XX	Human genome-derived single exon nucleic acid probes useful for analyzing	XX	27-SEP-2000; 2000US-0236359P.
XX	gene expression in human cervical epithelial cells.	XX	04-OCT-2000; 2000GB-00024263.
XX	Claim 25; SEQ ID NO 3004; 487pp; English.	XX	(MOLE-) MOLECULAR DYNAMICS INC.
XX	The present invention relates to a human single exon nucleic acid probes	XX	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	(SENP). The present sequence is one such probe. The SENPs are derived	XX	WPI; 2001-483447/52.
XX	from human Hela cells. The SENPs can be used to produce a single exon	XX	Human genome-derived single exon nucleic acid probes useful for analyzing
XX	microarray, which can be used for measuring human gene expression in a	XX	gene expression in human fetal liver.
XX	sample derived from human cervical epithelial cells. By measuring gene	XX	Claim 1; SEQ ID NO 3075; 639pp + Sequence Listing; English.
XX	expression, the probes are therefore useful in grading and/or staging of	XX	The invention relates to a single exon nucleic acid probe for measuring
XX	diseases of the cervix, notably cervical cancer. Note: The sequence data	XX	human gene expression in a sample derived from human foetal liver. The
XX	for this patent did not form part of the printed specification, but was	XX	single exon nucleic acid probes may be used for predicting, measuring and
XX	obtained in electronic format directly from WIPO at	XX	displaying gene expression in samples derived from human fetal liver. The
XX	ftp.wipo.int/pub/published_sequences	XX	present sequence is a single exon nucleic acid probe of the invention.
XX	Sequence 498 BP; 109 A; 140 C; 129 G; 120 T; 0 U; 0 Other;	XX	Note: The sequence data for this patent did not form part of the printed
XX	Query Match 13.0%; Score 314; DB 4; Length 498;	XX	specification, but was obtained in electronic format directly from WIPO
XX	Best Local Similarity 100.0%; Pred. No. 3.2e-143;	XX	at ftp.wipo.int/pub/published_sequences
XX	Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	XX	Sequence 498 BP; 109 A; 140 C; 129 G; 120 T; 0 U; 0 Other;
QY	1865 AGCTGAACCTCCAGCGTCTGGTCCCTATGCGCGCGAGTCTCTGAGGACTCTGACGAGG 1924	QY	1865 AGCTGAACCTCCAGCGTCTGGTCCCTATGCGCGCGAGTCTCTGAGGACTCTGACGAGG 1924
DB	498 AGCTGAACCTCCAGCGTCTGGTCCCTATGCGCGCGAGTCTCTGAGGACTCTGACGAGG 439	DB	498 AGCTGAACCTCCAGCGTCTGGTCCCTATGCGCGCGAGTCTCTGAGGACTCTGACGAGG 439
QY	1925 AGTCAAGGGGGTGGGCAAGAGATGGGATTTGGTACGATTGTGAGTCTCCACTCTCCCG 1984	QY	1925 AGTCAAGGGGGTGGGCAAGAGATGGGATTTGGTACGATTGTGAGTCTCCACTCTCCCG 1984
DB	438 AGTCAAGGGGGTGGGCAAGAGATGGGATTTGGTACGATTGTGAGTCTCCACTCTCCCG 379	DB	438 AGTCAAGGGGGTGGGCAAGAGATGGGATTTGGTACGATTGTGAGTCTCCACTCTCCCG 379
QY	1985 GCCAAGATGCCGAAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCCCTAA 2044	QY	1985 GCCAAGATGCCGAAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCCCTAA 2044
DB	378 GCCAAGATGCCGAAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCCCTAA 319	DB	378 GCCAAGATGCCGAAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCCCTAA 319
QY	2045 ACGGTGCTAATAGTGCAGACAGCGACAGTGCACCGAAAGAAACCGGCTTAGCGCTGATG 2104	QY	2045 ACGGTGCTAATAGTGCAGACAGCGACAGTGCACCGAAAGAAACCGGCTTAGCGCTGATG 2104
DB	318 ACGGTGCTAATAGTGCAGACAGCGACAGTGCACCGAAAGAAACCGGCTTAGCGCTGATG 259	DB	318 ACGGTGCTAATAGTGCAGACAGCGACAGTGCACCGAAAGAAACCGGCTTAGCGCTGATG 259
QY	2105 GTGCCAGCTGCCAAGGGCAGCTGCGCTGCACTCAGAAATCCCTTTGCTTAAGGCAACG 2164	QY	2105 GTGCCAGCTGCCAAGGGCAGCTGCGCTGCACTCAGAAATCCCTTTGCTTAAGGCAACG 2164
DB	258 GTGCCAGCTGCCAAGGGCAGCTGCGCTGCACTCAGAAATCCCTTTGCTTAAGGCAACG 199	DB	258 GTGCCAGCTGCCAAGGGCAGCTGCGCTGCACTCAGAAATCCCTTTGCTTAAGGCAACG 199
QY	2165 GTCTTCTCTGGAAG 2178	QY	2165 GTCTTCTCTGGAAG 2178
DB	198 GTCTTCTCTGGAAG 185	DB	198 GTCTTCTCTGGAAG 185
RESULT 23		RESULT 24	
ABAS4770/C		ABA44316/C	
ID ABA54770 standard; DNA; 498 BP.		ID ABA44316 standard; DNA; 498 BP.	
XX ABA54770;		XX ABA44316;	
XX AC		XX AC	
XX AC		XX AC	
DT 01-FEB-2002 (first entry)		DT 01-FEB-2002 (first entry)	
XX Human foetal liver single exon nucleic acid probe #3075.		XX Human breast cell single exon nucleic acid probe #3011.	
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.			
XX Homo sapiens.			
OS Homo sapiens.			
XX WO200157277-A2.			
PN		DE	

XX Human; microarray; single exon probe; gene expression; breast; disease;
KW cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157271-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000662.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes, useful
XX for measuring gene expression in sample derived from human breast,
XX comprises number of single exon nucleic acid probes.
XX
XX Claim 1; SEQ ID NO 3011; 327pp + Sequence Listing; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and Br 474 cells. The method involves contacting the
XX probes with a collection of detectably labelled nucleic acids derived
XX from mRNA of human breast, and then measuring the label bound to each
XX probe of the microarray. The probes are useful for verifying the
XX expression of regions of genomic DNA predicted to encode proteins. They
XX are useful for gene discovery, and for determining predisposition and/or
XX prognosing breast disease. Gene expression analysis is useful for
XX assessing the toxicity of chemical agents on cells. The microarray of
XX this invention presents a far greater diversity of probes for measuring
XX gene expression, with far less bias than expressed sequence tag
XX microarrays. The method is suitable for rapid production of functional
XX information from genomic sequence. The present sequence is a single exon
XX nucleic acid probe of the invention. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 498 BP; 109 A; 140 C; 129 G; 120 T; 0 U; 0 Other;
XX
XX Query Match 13.0%; Score 314; DB 4; Length 498;
XX Best Local Similarity 100.0%; Pred. No. 3.2e-143;
XX Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1865 AGCTGAACCTCAGCGTGTGGTGCCTATGCGCCGAGTCTCTGAGGACTCTGACGAGG 1924
XX
XX 498 AGCTGAACCTCAGCGTGTGGTGCCTATGCGCCGAGTCTCTGAGGACTCTGACGAGG 439
XX
XX 1925 AGTCAAAAGGGGCTGGGCAAGAGAGATGGGATGTGTACGATTGTGAGCTCCCACTCTCCG 1984
XX
XX 438 AGTCAAAAGGGGCTGGGCAAGAGAGATGGGATGTGTACGATTGTGAGCTCCCACTCTCCG 379
XX
XX 1985 GCCAAGATGCCGAGATGAGAGGCCACTCCGACAGGCTTCAAGAACCCATGACCCCTAA 2044
XX
XX 378 GCCAAGATGCCGAGATGAGAGGCCACTCCGACAGGCTTCAAGAACCCATGACCCCTAA 319
XX
XX 2045 ACGGTGCTAATAGTCAGACAGACGACGATGACCCGAAAGAAACGGCCTAGCGCTGATG 2104
XX
XX 318 ACGGTGCTAATAGTCAGACAGACGACGATGACCCGAAAGAAACGGCCTAGCGCTGATG 259

QY 2105 GTGCCAGCTGCCAAGGCCAGCGCTGCTGCACTCAGAAAATCCCTTTGCTAAGCAACG 2164
Dd |||||
258 GTGCCAGCTGCCAAGGCCAGCGCTGCTGCACTCAGAAAATCCCTTTGCTAAGCAACG 199
QY 2165 GTCTTCTCTGGAAG 2178
Dd |||||
198 GTCTTCTCTGGAAG 185
XX
XX RESULT 25
XX AAK28495/c
XX ID AAK28495 standard; DNA; 498 BP.
XX AC AAK28495;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 3052.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000668.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-489900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX
XX Example 4; SEQ ID NO 3052; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is one of
XX the probes of the invention
XX
XX Sequence 498 BP; 109 A; 140 C; 129 G; 120 T; 0 U; 0 Other;
XX
XX Query Match 13.0%; Score 314; DB 4; Length 498;
XX Best Local Similarity 100.0%; Pred. No. 3.2e-143;
XX Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1865 AGCTGAACCTCAGCGTGTGGTGCCTATGCGCCGAGTCTCTGAGGACTCTGACGAGG 1924
XX
XX 498 AGCTGAACCTCAGCGTGTGGTGCCTATGCGCCGAGTCTCTGAGGACTCTGACGAGG 439
XX
XX 1925 AGTCAAAAGGGGCTGGGCAAGAGAGATGGGATGTGTACGATTGTGAGCTCCCACTCTCCG 1984
XX
XX 438 AGTCAAAAGGGGCTGGGCAAGAGAGATGGGATGTGTACGATTGTGAGCTCCCACTCTCCG 379
XX
XX 1985 GCCAAGATGCCGAGATGAGAGGCCACTCCGACAGGCTTCAAGAACCCATGACCCCTAA 2044
XX
XX 378 GCCAAGATGCCGAGATGAGAGGCCACTCCGACAGGCTTCAAGAACCCATGACCCCTAA 319

1865 AGCTGAACCTCAGCGTGTCTGTGCGCCCTATGGCGCGAGTCTCTGAGGACTCTGACGAGG 1924
 438 AGCTGAACCTCAGCGTGTCTGTGCGCCCTATGGCGCGAGTCTCTGAGGACTCTGACGAGG 439
 1925 AGTCAAGGGGCTGGCGCAAGGAGATGGATTGGTACGATTGTGAGTCTCCCACTCTCCCG 1984
 438 AGTCAAGGGGCTGGCGCAAGGAGATGGATTGGTACGATTGTGAGTCTCCCACTCTCCCG 379
 1985 GCCAAGATGCCAAGATGAGGAGGCGCACTCCGCGAGGCTTCAAGAACCCATGACCTAA 2044
 378 GCCAAGATGCCAAGATGAGGAGGCGCACTCCGCGAGGCTTCAAGAACCCATGACCTAA 319
 2045 ACGGTGCTTAATAGTCAGACAGCGACAGTGAACCGAAGAAACCGCTAGGCTGTGATG 2104
 318 ACGGTGCTTAATAGTCAGACAGCGACAGTGAACCGAAGAAACCGCTAGGCTGTGATG 259
 2105 GTGCCAGCTGCCAAGCGCCAGCTGCCCTGCACTCAGAAATCCCTTGTAAAGGCAACG 2164
 258 GTGCCAGCTGCCAAGCGCCAGCTGCCCTGCACTCAGAAATCCCTTGTAAAGGCAACG 199

2165 GTCTTCTCTGGAAG 2178
 198 GTCTTCTCTGGAAG 185

RESULT 27
 ABS03008/c
 ID ABS03008 standard; DNA; 498 BP.
 XX
 AC ABS03008;
 XX
 XX
 19-AUG-2002 (first entry)
 XX
 DE Human genome-derived single exon probe from lung SEQ ID No 2999.
 XX
 KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease.
 XX
 OS Homo sapiens.
 XX
 XX WO200186003-A2.
 XX
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000665.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2002-114183/15.
 XX
 DR Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples.
 XX
 PS Claim 1; SEQ ID NO 2999; 634pp; English.
 CC The invention relates to a spatially-addressable set of single exon

2045 ACGGTGCTTAATAGTCAGACAGCGACAGTGAACCGAAGAAACCGCTAGGCTGTGATG 2104
 318 ACGGTGCTTAATAGTCAGACAGCGACAGTGAACCGAAGAAACCGCTAGGCTGTGATG 259
 2105 GTGCCAGCTGCCAAGCGCCAGCTGCCCTGCACTCAGAAATCCCTTGTAAAGGCAACG 2164
 258 GTGCCAGCTGCCAAGCGCCAGCTGCCCTGCACTCAGAAATCCCTTGTAAAGGCAACG 199

2165 GTCTTCTCTGGAAG 2178
 198 GTCTTCTCTGGAAG 185

RESULT 26
 AA102976/c
 ID AA102976 standard; DNA; 498 BP.
 XX
 AC AA102976;
 XX
 XX
 09-OCT-2001 (first entry)
 XX
 DE Probe #2967 used to measure gene expression in human breast sample.
 XX
 KW Probe; human; breast disease; breast cancer; development disorder; ss;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX
 OS Homo sapiens.
 XX
 XX WO200157270-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 29-JAN-2001; 2001WO-US000661.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-476286/51.
 XX
 DR Novel single exon nucleic acid probe used to measuring gene expression in
 XX a human breast.
 XX
 PF Claim 25; SEQ ID NO 2967; 322pp; English.
 XX
 CC The present invention relates to novel single exon nucleic acid probes.
 CC The present sequence is one such probe. The probes are useful for
 CC measuring human gene expression in a human breast sample, where the probe
 CC hybridises at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases
 CC include: breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocystic changes, proliferative breast disease and non-
 CC carcinoma tumours. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format
 CC directly from WIFO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 498 BP; 109 A; 140 C; 129 G; 120 T; 0 U; 0 Other;
 Query Match 13.0%; Score 314; DB 5; Length 498;
 Best Local Similarity 100.0%; Pred. No. 3.2e-143;
 Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

developmental or haematopoietic disorders. The nucleic acid can be used as probes, primers, in biological assays, to determine patterns of gene expression, to design ribozymes and to construct transgenic animals. The present sequence represents one of the 268 disclosed human G-protein-coupled protease cDNA sequences

XX	
SQ	Sequence 457 BP; 122 A; 113 C; 103 G; 119 T; 0 U; 0 Other;
Query Match	8.5%; Score 205; DB 6; Length 457;
Best Local Similarity	100.0%; Pred. No. 9.4e-90;
Matches 205; Conservative	0; Mismatches 0; Indels 0; Gaps 0;

QY 138 CTCCTGAGGAGTCTCACTTGGAGACATGGATGACAGTTCTGCCAGCTGGGGGCTCTGTCTC 197

Qy 198 TTCATTGAATGATGTCTCAAAATCACACACTTCTTTAGGACCAGTACCTGGTCTGTAGT 257

[illegible]

DB	223	T T A T T C G A G T T C A T C T G T A C C T G A T A A A C C A A A G G A T C A A G C C C T
QV	318	A G G T G A T G G C A T C G C T C C T C C A C A G

Db
283 AGGTGATGGCATGCTCTCTCCACAG 307

RESULT 30
AA115006/C
ID AA115006 standard; DNA; 489 BP.

XX
XX
AC
XX
AA115006;
- 2 Cam 2003 / 51-2003 011244.1

XX Probe #4939 for gene expression analysis in human cervical cell sample.
DE
XX

KW Probe; human; microarray; gene expression; cervical cancer; ss.
XX cervical cancer; ss.
XX cervical cancer

XX
PN
XX
WO200157278-A2.

PD	09-AUG-2001.
XX	
XX	
PF	30-JAN-2001; 2001WO-US000670.
...	

PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.

PR	03-AUG-2000	2000US-00042386.
PR	21-SEP-2000	2000US-0234687P.
PR	27-SEP-2000	2000US-0236359P.
PR	01-OCT-2000	2000CB-00024263

XX
XX
PA
XX
(MOLE-) MOLECULAR DYNAMICS INC.

FI FISH SG, HANDEL DR, CLASS W, KANSAS ST,
XX
DR WPI; 2001-488901/53.
XX

XX
gene expression in human cervical epithelial cells.
PT Human genome-derived single exon nucleic acid probes useful for analyzing

XX The present invention relates to human single exon nucleic acid probes
CC (cexp) the present sequence is one such probe. The SEMP are derived
aa

CC from human HeLa cells. The SNPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a

126 TCAGCTGGCAGCTCCGAGGCGAGTCTCACCTGGAGACATGGATGCAGGTTCTGCCAGCTG 185

121 TCAGCTGGCAGCTCCGAGGCGAGTCTCACCTGGAGACATGGATGCAGGTTCTGCCAGCTG 180

[illegible][illegible]

ABK30444
ID ABK30444 standard; cDNA; 457 BP.
XX

23-APR-2002 (first entry)

DE Human G-protein-coupled protease #214.
XX
KW Human; ss; gene; G-protein-coupled protease; gene therapy; transgenic;
Human; ss; gene; G-protein-coupled protease; gene therapy; transgenic;

KW	differentiative disorder; developmental disorder;
KW	haematopoietic disorder.
XX	

OS Homo sapiens.
XX
PN US6331427-B1.

AA
PD 18-DEC-2001.
XX
DF 26-MAR-1999: 99US-00280116.

XX
PR 26-MAR-1999; 99US-00280116.
XX

XX
PI Robison KE;
XX

DR WPI; 2002-129545/17.
XX
PT New polynucleotides encoding protease homologs of the G-protein-coupled
PT for identifying agonists and antagonists for

XX PT diagnosis and treatment of protease mediated disorders.
XX PS Disclosure: Col 243-244; 246pp; English.

XX The invention relates to an isolated human protease nucleic acid molecule
CC comprising a nucleotide sequence of 546 base pairs, one of 268 fully
CC complementary base pairs, also disclosed are production of an
CC

isolated polypeptide encoded by the nucleic acid, comprising introducing the nucleic acid into a host cell and culturing under conditions to express the protein from the nucleic acid, use of an antibody to detect

the encoded protein in a sample and to modulate its in vivo activity,
identifying agents that bind to the protein and identification of a
polynucleotide agent that modulates the expression of the nucleic acid or
its complement (i.e. gene therapy). The nucleic acid can be used to

CC identify an agent that modulates the expression or activity of the
CC nucleic acid, and can be used to isolate the protein. The nucleic acid
CC can be used in diagnostic assays for determining nucleic acid expression

as well as activity in the context of a biological sample (6,37,38), serum, cells, tissue) to determine whether an individual has a disease or disorder, or is at risk of developing a disease or disorder, associated with aberrant expression or activity of the nucleic acid. The nucleic

CC acid can be used to detect mutations in protease genes and gene
CC expression products such as mRNA. The nucleic acid can be used as
CC hybridisation probes to detect naturally-occurring genetic mutations in

CC protease gene. The nucleic acid can be used in drug screening methods.

CC identify agonists and antagonists that can be used to diagnose and treat protease mediated disorders e.g., proliferative, differentiative, and regenerative disorders.

CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging of
 CC diseases of the cervix, notably cervical cancer. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 489 BP; 162 A; 97 C; 94 G; 136 T; 0 U; 0 Other;
 Query Match 8.5%; Score 205; DB 4; Length 489;
 Best Local Similarity 100.0%; Pred. No. 9.3e-90;
 Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 312 AGCCCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTG 371
 Db 449 AGCCCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTG 390
 QY 372 TCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTG 431
 Db 389 TCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTG 330
 QY 432 TTTTGCCAATGCGAGCACTCAGTGTGTTAACTTACCTACACACCCTCTTGGCCAAATTACATGCT 491
 Db 329 TTTTGCCAATGCGAGCACTCAGTGTGTTAACTTACCTACACACCCTCTTGGCCAAATTACATGCT 270
 QY 492 ATCATGTAACACTCCAAAACATGT 516
 Db 269 ATCATGTAACACTCCAAAACATGT 245

RESULT 31
 ABA56739/c
 ID ABA56739 standard; DNA; 489 BP.
 XX AC ABA56739;
 XX DT 01-FEB-2002 (first entry)
 XX DE Human foetal liver single exon nucleic acid probe #5044.
 XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX OS Homo sapiens.
 XX XX
 XX PN WO200157277-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US0000669.
 XX PR 04-FEB-2000; 2000US-0180312P.
 XX PR 26-MAY-2000; 2000US-0207456P.
 XX PR 30-JUN-2000; 2000US-00608408.
 XX PR 03-AUG-2000; 2000US-00632366.
 XX PR 21-SEP-2000; 2000US-0234687P.
 XX PR 27-SEP-2000; 2000US-0236359P.
 XX PR 04-OCT-2000; 2000GB-00024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI; 2001-483447/52.
 XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
 XX PT gene expression in human fetal liver.
 XX PS Claim 1; SEQ ID NO 5044; 639pp + Sequence Listing; English.
 XX CC The invention relates to a single exon nucleic acid probe for measuring
 CC human gene expression in a sample derived from human foetal liver. The
 CC single exon nucleic acid probes may be used for predicting, measuring and
 CC displaying gene expression in samples derived from human fetal liver. The

CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 489 BP; 162 A; 97 C; 94 G; 136 T; 0 U; 0 Other;
 Query Match 8.5%; Score 205; DB 4; Length 489;
 Best Local Similarity 100.0%; Pred. No. 9.3e-90;
 Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 312 AGCCCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTG 371
 Db 449 AGCCCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTG 390
 QY 372 TCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTG 431
 Db 389 TCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTG 330
 QY 432 TTTTGCCAATGCGAGCACTCAGTGTGTTAACTTACCTACACACCCTCTTGGCCAAATTACATGCT 491
 Db 329 TTTTGCCAATGCGAGCACTCAGTGTGTTAACTTACCTACACACCCTCTTGGCCAAATTACATGCT 270
 QY 492 ATCATGTAACACTCCAAAACATGT 516
 Db 269 ATCATGTAACACTCCAAAACATGT 245

RESULT 32
 AAI36351/c
 ID AAI36351 standard; DNA; 489 BP.
 XX AC AAI36351;
 XX DT 17-OCT-2001 (first entry)
 XX DE Probe #5037 used to measure gene expression in human placenta sample.
 XX KW Probe; microarray; human; placenta; antenatal diagnosis;
 XX OS Genetic disorder; ss.
 XX XX
 XX OS Homo sapiens.
 XX XX
 XX PN WO200157272-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US0000663.
 XX PR 04-FEB-2000; 2000US-0180312P.
 XX PR 26-MAY-2000; 2000US-0207456P.
 XX PR 30-JUN-2000; 2000US-00608408.
 XX PR 03-AUG-2000; 2000US-00632366.
 XX PR 21-SEP-2000; 2000US-0234687P.
 XX PR 27-SEP-2000; 2000US-0236359P.
 XX PR 04-OCT-2000; 2000GB-00024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI; 2001-488897/53.
 XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
 XX PT gene expression in human placenta.
 XX PS Claim 25; SEQ ID NO 5037; 654pp; English.
 XX CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders

Example 4: SEQ ID NO 4956: 658bp + Sequence Listing; English.

XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression in
XX PT

```
PT a human breast.
XX Claim 25; SEQ ID NO 4765; 322pp; English.
PS
XX
XX The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and non-
CC carcinoma tumours. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 489 BP; 162 A; 97 C; 94 G; 136 T; 0 U; 0 Other;
SQ
Query Match 8.5%; Score 205; DB 5; Length 489;
Best Local Similarity 100.0%; Pred. No. 9.3e-90;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 312 AGCCCTAGGTGATGCGATCGCTCTCCACAGAAAGTTCTTTTCCCATCTCGAAGATTG 371
DB 449 AGCCCTAGGTGATGCGATCGCTCTCCACAGAAAGTTCTTTTCCCATCTCGAAGATTG 390
QY 372 TCTTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTG 431
DB 389 TCTTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTG 330
QY 432 TTTTGCCAAATCGAGCTGCGAGTGTAACTACACACCACTCTTCCCAATTACATGCT 491
DB 329 TTTTGCCAAATCGAGCTGCGAGTGTAACTACACACCACTCTTCCCAATTACATGCT 270
QY 492 ATCATGGAACACTCCAAACATGT 516
DB 269 ATCATGGAACACTCCAAACATGT 245
RESULT 37
ABS05002/C
ID ABS05002 standard; DNA; 489 BP.
AC
XX ABS05002;
XX
XX 19-AUG-2002 (first entry)
XX
XX Human genome-derived single exon probe from lung SEQ ID No 4993.
XX
XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenar syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX
XX Homo sapiens.
XX
XX WO2001186003-A2.
XX
XX 15-NOV-2001.
XX
XX 30-JAN-2001; 2001WO-US000665.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chan W, Rank DR;
XX
XX WPI; 2002-114183/15.
XX
XX Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples.
PT
XX Claim 1; SEQ ID NO 4993; 634pp; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human lung comprising single exon nucleic acid probes having one of
XX 12614 nucleic acid sequences mentioned in the specification, or their
XX complements or the 12387 open reading frames derived from the 12614
XX probes. Also included are a microarray comprising the novel set of probes
XX; the novel set of probes which hybridise at high stringency to a nucleic
XX acid expressed in the human lung; measuring gene expression in a sample
XX derived from human lung, comprising (a) contacting the array with a
XX collection of detectably labeled nucleic acids derived from human lung
XX mRNA, and (b) measuring the label detectably bound to each probe of the
XX array; identifying exons in a eukaryotic genome, comprising (a)
XX algorithmically predicting at least one exon from genomic sequences of
XX the eukaryote; and (b) detecting specific hybridisation of detectably
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
XX having a fragment identical to the predicted exon, the probe is included
XX in the above mentioned microarray; assigning exons to a single gene,
XX comprising (a) identifying exons from genomic sequence by the method
XX above and (b) measuring the expression of each of the exons in several
XX tissues and/or cell types using hybridisation to a single exon
XX microarrays having a probe with the exon, where a common pattern of
XX expression of the exons in the tissues and/or cell types indicates that
XX the exons should be assigned to a single gene; a peptide comprising one
XX of 12011 sequences, mentioned in the specification, or encoded by the
XX probes/open reading frames (ORF). The probes are used for gene expression
XX analysis, and for identifying exons in a gene, particularly using human
XX lung derived mRNA and for the study of lung diseases such as asthma, lung
XX cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
XX disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
XX tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
XX Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
XX histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
XX Karagenar syndrome, fibrocystic pulmonary dysplasia, primary ciliary
XX dyskinesia, pulmonary hypertension and hyaline membrane disease. The
XX present sequence is a single exon probe of the invention. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 489 BP; 162 A; 97 C; 94 G; 136 T; 0 U; 0 Other;
SQ
Query Match 8.5%; Score 205; DB 6; Length 489;
Best Local Similarity 100.0%; Pred. No. 9.3e-90;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 312 AGCCCTAGGTGATGCGATCGCTCTCCACAGAAAGTTCTTTTCCCATCTCGAAGATTG 371
DB 449 AGCCCTAGGTGATGCGATCGCTCTCCACAGAAAGTTCTTTTCCCATCTCGAAGATTG 390
QY 372 TCTTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTG 431
DB 389 TCTTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTG 330
QY 432 TTTTGCCAAATCGAGCTGCGAGTGTAACTACACACCACTCTTCCCAATTACATGCT 491
DB 329 TTTTGCCAAATCGAGCTGCGAGTGTAACTACACACCACTCTTCCCAATTACATGCT 270
QY 492 ATCATGGAACACTCCAAACATGT 516
DB 269 ATCATGGAACACTCCAAACATGT 245
```


Db 269 ATCATGTAACACTCCAAACATGT 245

RESULT 38
AAI24193/c
ID AAI24193 standard; DNA; 198 BP.
AC AAI24193;
XX
XX 12-OCT-2001 (first entry)
DE Probe #14126 for gene expression analysis in human cervical cell sample.
XX Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX Homo sapiens.
XX WO200157278-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000670.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-489901/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human cervical epithelial cells.
XX Claim 25; SEQ ID NO 14126; 487pp; English.
XX The present invention relates to human single exon nucleic acid probes
XX (SENP). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging of
XX diseases of the cervix, notably cervical cancer. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 198 BP; 53 A; 37 C; 53 G; 55 T; 0 U; 0 Other;
Query Match 8.2%; Score 198; DB 4; Length 198;
Best Local Similarity 100.0%; Pred. No. 2.6e-86;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 314 CCTAGGTGATGCATCGCTCTCCACAGAAAGTTCTTTTCCCATCTCGAAGATTGTC 373
Db 198 CCCTAGGTGATGCATCGCTCTCCACAGAAAGTTCTTTTCCCATCTCGAAGATTGTC 139
QY 374 TTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTGGGCAATACCTGTT 433
Db 138 TTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTGGGCAATACCTGTT 79
QY 434 TTGCCAATGCAGCTGAGTGTCTTAACCTACACACCACTCTTGGCCAAATTACATGCTAT 493
Db 78 TTGCCAATGCAGCTGAGTGTCTTAACCTACACACCACTCTTGGCCAAATTACATGCTAT 19
QY 494 CACATGAACACTCCAAAA 511
Db 18 CACATGAACACTCCAAAA 1

Db 269 ATCATGTAACACTCCAAACATGT 245

RESULT 38
AAI24193/c
ID AAI24193 standard; DNA; 198 BP.
AC AAI24193;
XX
XX 12-OCT-2001 (first entry)
DE Probe #14126 for gene expression analysis in human cervical cell sample.
XX Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX Homo sapiens.
XX WO200157278-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000670.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-489901/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human cervical epithelial cells.
XX Claim 25; SEQ ID NO 14126; 487pp; English.
XX The present invention relates to human single exon nucleic acid probes
XX (SENP). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging of
XX diseases of the cervix, notably cervical cancer. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 198 BP; 53 A; 37 C; 53 G; 55 T; 0 U; 0 Other;
Query Match 8.2%; Score 198; DB 4; Length 198;
Best Local Similarity 100.0%; Pred. No. 2.6e-86;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 314 CCTAGGTGATGCATCGCTCTCCACAGAAAGTTCTTTTCCCATCTCGAAGATTGTC 373
Db 198 CCCTAGGTGATGCATCGCTCTCCACAGAAAGTTCTTTTCCCATCTCGAAGATTGTC 139
QY 374 TTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTGGGCAATACCTGTT 433
Db 138 TTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTGGGCAATACCTGTT 79
QY 434 TTGCCAATGCAGCTGAGTGTCTTAACCTACACACCACTCTTGGCCAAATTACATGCTAT 493
Db 78 TTGCCAATGCAGCTGAGTGTCTTAACCTACACACCACTCTTGGCCAAATTACATGCTAT 19
QY 494 CACATGAACACTCCAAAA 511
Db 18 CACATGAACACTCCAAAA 1

ID ABA51313 standard; DNA; 198 BP.
 XX
 AC ABA51313;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human breast cell single exon nucleic acid probe #10008.
 XX
 KW Human; microarray; single exon probe; gene expression; breast; disease;
 KW cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157271-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US003662.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-496933/54.
 XX
 PT New spatially-addressable set of single exon nucleic acid probes, useful
 PT for measuring gene expression in sample derived from human breast,
 PT comprises number of single exon nucleic acid probes.
 XX
 PS Claim 4; SEQ ID NO 10008; 327pp + Sequence Listing; English.
 CC
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and Br 474 cells. The method involves contacting the
 CC probes with a collection of detectably labelled nucleic acids derived
 CC from mRNA of human breast, and then measuring the label bound to each
 CC probe of the microarray. The probes are useful for verifying the
 CC expression of regions of genomic DNA predicted to encode proteins. They
 CC are useful for gene discovery, and for determining predisposition and/or
 CC prognosing breast disease. Gene expression analysis is useful for
 CC assessing the toxicity of chemical agents on cells. The microarray of
 CC this invention presents a far greater diversity of probes for measuring
 CC gene expression, with far less bias than expressed sequence tag
 CC microarrays. The method is suitable for rapid production of functional
 CC information from genomic sequence. The present sequence is a single exon
 CC nucleic acid probe of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 198 BP; 53 A; 37 C; 53 G; 55 T; 0 U; 0 Other;
 Query Match 8.2%; Score 198; DB 4; Length 198;
 Best Local Similarity 100.0%; Pred. No. 2.6e-86;
 Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 314 CCCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTGTC 373
 DB 198 CCCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTGTC 139
 QY 374 TTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTTGGCAATACCTGTT 433
 DB 138 TTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTTGGCAATACCTGTT 79
 QY 434 TTGCCAATGCAGCTGAGTGTAACTTAACTACACACCACCTCTTCCCAATTACATGCTAT 493

RESULT 40
 AAI49479/c
 ID AAI49479 standard; DNA; 198 BP.
 XX
 AC AAI49479;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #18165 used to measure gene expression in human placenta sample.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000663.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human placenta.
 XX
 PS Claim 25; SEQ ID NO 18165; 654pp; English.
 CC
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders
 XX
 SQ Sequence 198 BP; 53 A; 37 C; 53 G; 55 T; 0 U; 0 Other;
 Query Match 8.2%; Score 198; DB 4; Length 198;
 Best Local Similarity 100.0%; Pred. No. 2.6e-86;
 Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 314 CCCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTGTC 373
 DB 198 CCCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTGTC 139
 QY 374 TTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTTGGCAATACCTGTT 433
 DB 138 TTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTTGGCAATACCTGTT 79
 QY 434 TTGCCAATGCAGCTGAGTGTAACTTAACTACACACCACCTCTTCCCAATTACATGCTAT 493
 DB 78 TTGCCAATGCAGCTGAGTGTAACTTAACTACACACCACCTCTTCCCAATTACATGCTAT 19
 QY 494 CACATGAACACTCCAAA 511
 DB 18 CACATGAACACTCCAAA 1
 RESULT 41
 ABA51313/c

```
Db 78 TTGCAATGCAGCTGCAGTGTTTAACTTACATACACACCTCTTGCCCAATTACATGCTAT 19
Qy 494 CACATGAACACTCCAAAA 511
Db 18 CACATGAACACTCCAAAA 1

RESULT 42
AAK43414/c
ID AAK43414 standard; DNA; 198 BP.
XX
AC AAK43414;
XX
DT 06-NOV-2001 (first entry)
DE
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 17971.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000668.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
PS Example 4; SEQ ID NO 17971; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention
XX
SQ Sequence 198 BP; 53 A; 37 C; 53 G; 55 T; 0 U; 0 Other;

Query Match
Best Local Similarity 8.2%; Score 198; DB 4; Length 198;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 314 CCCTAGGTGATGGCATCGCTCCTCCACAGAAAGTTCTTTCCCATCTCAGAGATTGTC 373
Db 198 CCCTAGGTGATGGCATCGCTCCTCCACAGAAAGTTCTTTCCCATCTCAGAGATTGTC 139

Qy 374 TTAAGTGGCAACAACCTCATAGATTGGAGCTGGGCTCCAGAAATTGGGCAATACCTGTT 433
Db 138 TTAAGTGGCAACAACCTCATAGATTGGAGCTGGGCTCCAGAAATTGGGCAATACCTGTT 79

Qy 434 TTGCAATGCAGCTGCAGTGTTTAACTTACATACACACCTCTTGCCCAATTACATGCTAT 493
Db 78 TTGCAATGCAGCTGCAGTGTTTAACTTACATACACACCTCTTGCCCAATTACATGCTAT 19
```

```
Qy 494 CACATGAACACTCCAAAA 511
Db 18 CACATGAACACTCCAAAA 1

RESULT 43
ABS43035/c
ID ABS43035 standard; DNA; 198 BP.
XX
AC ABS43035;
XX
DT 25-FEB-2003 (first entry)
DE
XX
DE Human liver single exon probe, SEQ ID No 18025.
XX
KW Human; single exon nucleic acid probe; liver; cirrhosis;
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW coronary heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157273-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000664.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488998/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
PS Claim 4; SEQ ID NO 18025; 658pp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENp) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABS25011-ABS51005 represent human
CC liver single exon nucleic acid probes of the invention. Note: The
CC sequence information for this patent does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 198 BP; 53 A; 37 C; 53 G; 55 T; 0 U; 0 Other;

Query Match
Best Local Similarity 8.2%; Score 198; DB 4; Length 198;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 314 CCCTAGGTGATGGCATCGCTCCTCCACAGAAAGTTCTTTCCCATCTCAGAGATTGTC 373
Db 198 CCCTAGGTGATGGCATCGCTCCTCCACAGAAAGTTCTTTCCCATCTCAGAGATTGTC 139

Qy 374 TTAAGTGGCAACAACCTCATAGATTGGAGCTGGGCTCCAGAAATTGGGCAATACCTGTT 433
```

Db 138 TTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGCAATACCTGTT 79
QY 434 TTGCAATGCGACACTGCGAGTGTTTAACTACACACCACTCTTCCCAATTACATGCTAT 493
Db 78 TTGCAATGCGACACTGCGAGTGTTTAACTACACACCACTCTTCCCAATTACATGCTAT 19
QY 494 CACATGAACACTCCAAA 511
Db 18 CACATGAACACTCCAAA 1

RESULT 44
AAI09758/c
ID AAI09758 standard; DNA; 198 BP.
XX AAI09758;
AC AAI09758;
XX 09-OCT-2001 (first entry)
DT Probe #9749 used to measure gene expression in human breast sample.
DE Probe; human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
KW Homo sapiens.
OS WO200157270-A2.
PN XX
XX 09-AUG-2001.
PD XX
XX 29-JAN-2001; 2001WO-US000661.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX Novel single exon nucleic acid probe used to measuring gene expression in
PT a human breast.
PS Claim 25; SEQ ID NO 9749; 322pp; English.
XX The present invention relates to novel single exon nucleic acid probes.
XX The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and non-
CC carcinoma tumours. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 198 BP; 53 A; 37 C; 53 G; 55 T; 0 U; 0 Other;
SQ

Query Match 8.2%; Score 198; DB 5; Length 198;
Best Local Similarity 100.0%; Pred. No. 2.6e-86;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 CCTAGGTGATGCGATCGCTCTCCACAGAAAGTCTTTTCCCATCTGAGAGATTGTC 373
Db 198 CCTAGGTGATGCGATCGCTCTCCACAGAAAGTCTTTTCCCATCTGAGAGATTGTC 139

QY 374 TTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGCAATACCTGTT 433
Db 138 TTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGCAATACCTGTT 79
QY 434 TTGCAATGCGACACTGCGAGTGTTTAACTACACACCACTCTTCCCAATTACATGCTAT 493
Db 78 TTGCAATGCGACACTGCGAGTGTTTAACTACACACCACTCTTCCCAATTACATGCTAT 19
QY 494 CACATGAACACTCCAAA 511
Db 18 CACATGAACACTCCAAA 1

RESULT 45
ABS17512/c
ID ABS17512 standard; DNA; 198 BP.
XX ABS17512;
AC ABS17512;
XX 19-AUG-2002 (first entry)
DT Human genome-derived single exon probe ORF from lung SEQ ID No 17503.
DE Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Rudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenen syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease; open reading frame; ORF.
OS Homo sapiens.
XX WO200186003-A2.
XX 15-NOV-2001.
XX 30-JAN-2001; 2001WO-US000665.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2002-114183/15.
XX Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples.
XX Claim 4; SEQ ID NO 17503; 634pp; English.
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of probes
CC; the novel set of probes which hybridise at high stringency to a nucleic
CC acid expressed in the human lung; measuring gene expression in a sample
CC derived from human lung, comprising (a) contacting the array with a
CC collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of the
CC array; identifying exons in a eukaryotic genome, comprising (a)

algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karagenar syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a single exon probe open reading frame of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Query Match
Best Local Similarity 8.2%; Score 198; DB 6; Length 198;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 314 CCTAGTGTGTCGATCGTCTCTCCACAGAAAGTCTTTTCCATCTGAGAAGATTGTC 373
Db 198 CCTAGTGTGTCGATCGTCTCTCCACAGAAAGTCTTTTCCATCTGAGAAGATTGTC 139
QY 374 TTAAGTGGCAACAACTCATAGAGTTGGAGCTGGCTCCAGAAATTGGGCAATACCTGTT 433
Db 138 TTAAGTGGCAACAACTCATAGAGTTGGAGCTGGCTCCAGAAATTGGGCAATACCTGTT 79
QY 434 TTGCCAATGCAGCACTGAGTGTAACTACACACACCCTCTTGCCAAATTACATGCTAT 493
Db 78 TTGCCAATGCAGCACTGAGTGTAACTACACACACCCTCTTGCCAAATTACATGCTAT 19
QY 494 CACATGACACTCCAAA 511
Db 18 CACATGACACTCCAAA 1

RESULT 46
ABQ95217
ID ABQ95217 standard; DNA; 213 BP.
XX
AC ABQ95217;
XX
DT 28-OCT-2002 (first entry)
XX
DE Tumour suppression-related oligonucleotide #868.
XX
KW Tumour; cytostatic; antiviral; neuroprotective; nootropic; neuroleptic;
KW tumour suppression; tumour reversion; apoptosis; viral resistance; human;
KW viral infection; cell degeneration disease; neurodegeneration; ds;
KW Alzheimer's disease; schizophrenia; immune disease; inflammatory disease.
OS Homo sapiens.
XX
FN FR2819824-A1.
XX
PD 26-JUL-2002.
XX
FF 23-JAN-2001; 2001FR-00000899.

23-JAN-2001; 2001FR-00000899.
(MOLE-) MOLECULAR ENGINES LAB SA.
Teletman A, Anson R, Tuijnder M, Susini L;
WPI; 2002-610803/66.
New nucleic acid implicated e.g. in tumor suppression, useful for diagnosis of tumors, viral infection and cellular degeneration and for drug screening.
Claim 1; Page 256; 623pp; French.
The present invention relates to novel human nucleic acid sequences (I). The present sequence is one such nucleic acid sequence. Expression of (I) are implicated in tumour suppression or reversion and apoptosis and viral resistance. (I) are useful as probes or primers for detecting, identifying, measuring and/or amplifying nucleic acid sequences, as antisense reagents and for recombinant production of polypeptides. (I), polypeptides (II) encoded by (I), vector containing (I), cells containing these vectors and antibodies (Ab) against (II) are all useful for treatment/prevention of viral, tumour and cell degeneration diseases (especially neurodegeneration, such as Alzheimer's disease and schizophrenia). Analysing the expression of (I) is also useful for diagnosis and/or prognosis of such diseases. Transgenic animals carrying (I) are used for studying the aetiology of these diseases (also immune and inflammatory diseases). Note: In the present specification, SEQ ID 1 to 2280 are claimed in Claim 1, however only SEQ ID 1 to 2270 are shown in the specification
Sequence 213 BP; 64 A; 44 C; 47 G; 58 T; 0 U; 0 Other;
Query Match 5.7%; Score 137; DB 6; Length 213;
Best Local Similarity 100.0%; Pred. No. 2.2e-56;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 918 GCTTGATGGAGAAAACCTCGTACAGTCGAGCAAGTGTAAAAAGATGGTTCCAGTTCAA 977
Db 70 GCTTGATGGAGAAAACCTCGTACAGTCGAGCAAGTGTAAAAAGATGGTTCCAGTTCAA 129
QY 978 GAGGTTCACTATCCATAGATCTCTAAATGCTTTACACTTCTCTGAAACGTTTGCAAA 1037
Db 130 GAGGTTCACTATCCATAGATCTCTAAATGCTTTACACTTCTCTGAAACGTTTGCAAA 189
QY 1038 TTTTACCGGTGAAAAA 1054
Db 190 TTTTACCGGTGAAAAA 206
RESULT 47
ABN62290/c
ID ABN62290 standard; cDNA; 292 BP.
XX
AC ABN62290;
XX
DT 28-JUN-2002 (first entry)
XX
DE Human cancer related polynucleotide SEQ ID NO 2257.
XX
KW Human; cytostatic; gene expression; gene mapping; tissue profiling;
KW gene therapy; cancer; tumour; gene; ss.
OS Homo sapiens.
XX
FN WO200214500-A2.
XX
PD 21-FEB-2002.
XX
FF 16-AUG-2001; 2001WO-US025840.
XX
PR 16-AUG-2000; 2000US-0226326P.

```
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
PI Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
XX
XX WPI; 2002-241905/29.
XX
XX New nucleic acid for producing a polypeptide, detecting differentially
XX expressed genes correlated with a cancerous state of a mammalian cell,
XX and inhibiting tumor growth.
XX
XX Claim 1; SEQ ID NO 2257; 883pp + Sequence Listing; English.
XX
XX The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
XX with cytoskeletal activity. The polynucleotide is used to produce a
XX polypeptide, to detect differentially expressed genes correlated with a
XX cancerous state of a mammalian cell and to inhibit tumour growth. The
XX polynucleotide is used as a probe in mapping and tissue profiling. The
XX encoded polypeptide and antibodies to the polypeptide can also be used
XX for therapeutic and diagnostic purposes. The polynucleotide is useful for
XX gene therapy. Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 292 BP; 103 A; 65 C; 62 G; 62 T; 0 U; 0 Other;
XX
XX Query Match 4.1%; Score 100; DB 6; Length 292;
XX Best Local Similarity 99.3%; Pred. No. 3.1e-38;
XX Matches 150; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 312 AGCCCTAGGTGGGATCGCTCTCCACAGAAAGTCTTTTCCCATCTCAGAGATTG 371
Db |||||
QY 151 AGCCCTAGGTGGGATCGCTCTCCACAGAAAGTCTTTTCCCATCTCAGAGATTG 92
Db |||||
QY 372 TCTTAGTGGCACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGCAATACCTG 431
Db |||||
QY 91 TCTTAGTGGCACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGCAATACCTG 32
Db |||||
QY 432 TTTTGGCAATGCAGCACTGCAGTGTTTAACC 462
Db 31 TTTTGGCAATGCAGCACTGCAGTGTTTAACC 1
XX
RESULT 48
AAK57307
ID AAK57307 standard; cDNA; 657 BP.
XX
AC AAK57307;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:2367.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytoskeletal; gene therapy; vaccine; metastasis; ss.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US0001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX
XX 04-FEB-2000; 2000US-0180628P.
XX
XX 24-FEB-2000; 2000US-0184664P.
XX
XX 02-MAR-2000; 2000US-0186350P.
XX
XX 16-MAR-2000; 2000US-0189874P.
XX
XX 17-MAR-2000; 2000US-0190076P.
XX
XX 18-APR-2000; 2000US-0198123P.
XX
XX 19-MAY-2000; 2000US-0205515P.
XX
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214386P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216547P.
PR 07-JUL-2000; 2000US-0216380P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220363P.
PR 26-JUL-2000; 2000US-0220364P.
PR 14-AUG-2000; 2000US-0224318P.
PR 14-AUG-2000; 2000US-0224319P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229487P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
```

PR	20-OCT-2000;	2000US-024117855P
PR	20-OCT-2000;	2000US-02411786P
PR	20-OCT-2000;	2000US-02411787P
PR	20-OCT-2000;	2000US-02411808P
PR	20-OCT-2000;	2000US-0241808P
PR	20-OCT-2000;	2000US-0241826P
PR	01-NOV-2000;	2000US-0244617P
PR	08-NOV-2000;	2000US-0246474P
PR	08-NOV-2000;	2000US-0246475P
PR	08-NOV-2000;	2000US-0246476P
PR	08-NOV-2000;	2000US-0246477P
PR	08-NOV-2000;	2000US-0246478P
PR	08-NOV-2000;	2000US-0245323P
PR	08-NOV-2000;	2000US-0245324P
PR	08-NOV-2000;	2000US-0245325P
PR	08-NOV-2000;	2000US-0245326P
PR	08-NOV-2000;	2000US-0245327P
PR	08-NOV-2000;	2000US-0245328P
PR	08-NOV-2000;	2000US-0245329P
PR	08-NOV-2000;	2000US-024610P
PR	08-NOV-2000;	2000US-024611P
PR	08-NOV-2000;	2000US-0246613P
PR	17-NOV-2000;	2000US-0249207P
PR	17-NOV-2000;	2000US-0249208P
PR	17-NOV-2000;	2000US-0249209P
PR	17-NOV-2000;	2000US-0249211P
PR	17-NOV-2000;	2000US-0249212P
PR	17-NOV-2000;	2000US-0249213P
PR	17-NOV-2000;	2000US-0249214P
PR	17-NOV-2000;	2000US-0249215P
PR	17-NOV-2000;	2000US-0249216P
PR	17-NOV-2000;	2000US-0249217P
PR	17-NOV-2000;	2000US-0249218P
PR	17-NOV-2000;	2000US-0249244P
PR	17-NOV-2000;	2000US-0249245P
PR	17-NOV-2000;	2000US-0249246P
PR	17-NOV-2000;	2000US-0249255P
PR	17-NOV-2000;	2000US-0249297P
PR	17-NOV-2000;	2000US-0249299P
PR	17-NOV-2000;	2000US-0249300P
PR	01-DEC-2000;	2000US-0250160P
PR	01-DEC-2000;	2000US-0250391P
PR	05-DEC-2000;	2000US-0251030P
PR	05-DEC-2000;	2000US-0251988P
PR	05-DEC-2000;	2000US-0256719P
PR	06-DEC-2000;	2000US-0251479P
PR	08-DEC-2000;	2000US-0251856P
PR	08-DEC-2000;	2000US-0251868P
PR	08-DEC-2000;	2000US-0251869P
PR	08-DEC-2000;	2000US-0251989P
PR	08-DEC-2000;	2000US-0251990P
PR	11-DEC-2000;	2000US-0254097P
PR	05-JAN-2001;	2001US-0259678P

example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patients own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/haematopoietic-related diseases, especially cancers and cancer metastases of haematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/haematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAK82169 represent sequences used in the exemplification of the present invention

Sequence 657 BP; 180 A; 152 C; 163 G; 146 T; 0 U; 16 Other;

XX Lal P, Yue H, Tang YT, Lu DAM, Azimzai Y, Au-Young J;
 PI Hillman JL, Baughn MR, Yao MG, Burford N, Batra S, Policky JJ;
 XX WPI; 2002-034502/04.
 DR P-PSDB; AAE15241.
 XX New human RNA metabolism protein for diagnosing or treating nervous
 PT system disorders, autoimmune/inflammatory disorders, cell proliferative
 FT disorders and developmental disorders.
 XX Claim 11; Page 172; 196pp; English.
 XX The invention relates to human RNA metabolism proteins (RMEP) and their
 CC corresponding cDNA molecules. RMEP and its DNA are used for diagnosing,
 CC treating and preventing nervous system disorders (epilepsy, dementia,
 CC stroke, Alzheimer's disease, Huntington's disease, Parkinson's disease);
 CC prion diseases; fatal familial insomnia, nutritional and metabolic
 CC diseases of the nervous system; inherited, metabolic, endocrine and toxic
 CC myopathy; mental disorders (mood, anxiety, schizophrenic disorders) (AIDS-
 CC amnesia and Tourette's disorder; autoimmune/inflammatory disorders (AIDS-
 CC acquired immune deficiency syndrome, allergies, anaemia, asthma, gout,
 CC atherosclerosis, Crohn's disease, diabetes mellitus, glomerulonephritis,
 CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid
 CC arthritis, osteoporosis, pancreatitis, systemic lupus erythematosus,
 CC ulcerative colitis, and infections); cell proliferative disorders (cancer
 CC arteriosclerosis, cirrhosis, hepatitis, psoriasis); and developmental
 CC disorders (renal tubular acidosis). RMEP DNA is useful in drug screening
 CC techniques, gene therapy and for creating transgenic animals. The present
 CC sequence is human RMEP-4 cDNA
 XX
 SQ Sequence 637 BP; 175 A; 130 C; 150 G; 182 T; 0 U; 0 Other;
 Query Match 1.2%; Score 30; DB 6; Length 637;
 Best Local Similarity 100.0%; Pred. No. 0.00065;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
 DB 37 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 8
 RESULT 50
 AAD08062
 ID AAD08062 standard; cDNA; 733 BP.
 AC AAD08062;
 XX
 XX 06-AUG-2001 (first entry)
 DT Human extracellular matrix and cell adhesion molecule-18 (XNAD-18) cDNA.
 DE Human; extracellular matrix and cell adhesion molecule; XNAD;
 KW Gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia;
 KW Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dystrophy;
 KW sickle cell anaemia; thalassaemia; autoimmune disorder; adenocarcinoma;
 KW inflammatory disorder; acquired immune deficiency syndrome; AIDS;
 KW Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma;
 KW Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer;
 KW glomerulonephritis; multiple sclerosis; Grave's disease; osteoarthritis;
 KW osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis;
 KW infection; cell proliferative disorder; actinic keratosis; myeloma;
 KW arteriosclerosis; neotropic; anticonvulsant; antithyroid; nephrotropic;
 KW neuroprotective; dermatological; chromosome 22; ss.
 XX Homo sapiens.
 OS
 XX Location/Qualifiers
 Key 146..703
 FH CDS
 FT /*tag= a
 FT /product= "Human extracellular matrix and cell adhesion
 FT molecule (XNAD)"
 XX

FN WO200142285-A2.
 XX 14-JUN-2001.
 PD 05-DEC-2000; 2000WO-US032990.
 XX 10-DEC-1999; 99US-0172852P.
 PR 16-DEC-1999; 99US-0172354P.
 XX (INCY-) INCYTE GENOMICS INC.
 PA Yue H, Tang YT, Lal P, Burford N, Azimzai Y, Patterson C;
 PI Baughn MR, Lu DAM, Shah P, Au-Young J;
 XX WPI: 2001-381632/40.
 DR P-PSDB; AAE03854.
 XX New human extracellular matrix and cell adhesion molecules and
 PT polynucleotide sequences encoding them, useful for diagnosis, prevention,
 FT treatment of genetic, autoimmune and cell proliferative disorders.
 XX Claim 5; Page 131; 135pp; English.
 XX The present cDNA sequence encodes human extracellular matrix and cell
 CC adhesion molecule (XNAD). The XNAD is used for screening a compound for
 CC effectiveness as an agonist or antagonist of XNAD. The identified agonist
 CC or antagonist are used for treating a disease or condition associated
 CC with decreased or increased expression of functional XNAD. The
 CC polynucleotides encoding XNAD are useful in somatic or germline gene
 CC therapy to correct a genetic deficiency, to express a conditionally
 CC lethal gene product and to express a protein which affords protection
 CC against intracellular parasites and also for diagnosis of disorders
 CC associated with expression of XNAD. They are also used for generating
 CC hybridisation probes useful in mapping the naturally occurring genomic
 CC sequences and to create knock in humanised animals (pigs) or transgenic
 CC animals (mice or rats) to model human diseases. Oligonucleotide or longer
 CC fragments derived from the polynucleotide sequences may be used as
 CC elements on a microarray. Antibodies which specifically bind XNAD may be
 CC used for the diagnosis of disorders associated with the expression of
 CC XNAD, or in assays to monitor patients being treated with XNAD. Diseases
 CC diagnosed, prevented or treated include genetic disorders such as
 CC adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's
 CC disease, myotonic dystrophy, sickle cell anaemia, thalassaemia,
 CC autoimmune/inflammatory disorders such as acquired immune deficiency
 CC syndrome (AIDS), Addison's disease, diabetes mellitus, atopic dermatitis,
 CC atherosclerosis, Crohn's disease, diabetes mellitus, ulcerative colitis,
 CC glomerulonephritis, multiple sclerosis, Grave's disease, osteoarthritis,
 CC osteoporosis, psoriasis, rheumatoid arthritis, ulcerative colitis,
 CC bacterial, fungal, parasitic, protozoal and helminthic infections and
 CC cell proliferative disorders such as actinic keratosis, arteriosclerosis
 CC and cancer including breast, bladder, bone marrow, brain and uterus
 CC cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma
 XX
 SQ Sequence 733 BP; 107 A; 240 C; 245 G; 141 T; 0 U; 0 Other;
 Query Match 1.2%; Score 30; DB 4; Length 733;
 Best Local Similarity 100.0%; Pred. No. 0.00065;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
 DB 696 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 725
 Search completed: August 17, 2004, 00:36:39
 Job time : 1367 secs

Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 4596.f. For more information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DF001AD12QP2&cluster=4596.f>. Contact : Feng Liang Email : liang@lifetech.com URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 <http://www.fulllength.invitrogen.com/> ID : CS0DF001AD12QP2.

FEATURES
source

```

1. 1201
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="CSODF001YG23"
    /tissue_type="FETAL BRAIN"
    /dev_stage="fetal"
    /clone_lib="Homo sapiens FETAL BRAIN"
    /note="Organ: brain; Vector: pCMWSORT 6; 1st strand cDNA
    was primed with a NotI-oligo(dN) primer. Five prime end
    was enriched, double-strand cDNA was digested with Not I and
    cloned into the Not I and EcoRV sites of the pCMWSORT 6
    vector library was not normalized."

```

ORIGIN

```
Query Match      37.7%; Score 909; DB 13; Length 1201;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 959; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

23 CGCGCGCGAGGGGATGAGCGAGCGCGCGCGGTACAGTTGAACAATGACCATAG 82

b6 CGGGGCGCAGGGGAAGGAGCGACGCCCCGAAATTCCTTCAAGAATTCTGTACTTTTG
83 TTGCACAAGCTTCTGAATCTTCAGACCCTCATCACCTTATCAGAAATCAGCTGGCAGCTCCG

y y

126 TTGACAAAGCTTCTGAATCTTTCAGACCCATCAGCCTATTCAGAAATCAGCCTGGCAGCTCG 185

143 AGGAGTCTCACTGGAGACATGGATGCCAGTTCGCCAGCTGGGGTGCTGTCTTCTTCAAT 202

186 AGGCGTCTTACCTCGAGACATGGATGCAGGTTCTGCCAGCTGGGGTCTGTCTTCTCAT 245

203 TGAATGATGTGTCATCATCACACATTTCTTTAGGACACAGTACCTGGTCTGAGAGTAAAT 305
 246 TGAATGATGTGTCATCATCACACATTTCTTTAGGACACAGTACCTGGTCTGAGTATT 305

263 CGAGTTCATCTGTACTGTATAATCAAAACCATCACCAAAAGGATCAAGCCCTAGGTG 322

306 CGAGTTCATCTGTAACGCATTACTTTGGTG
323 ATGGCATCGTCTCCTCACAGAAAGTTCTTTCCCATCTGAGAGAGATTGTCTTAAATGGC 382

QY

366 ATGGCATCGCTCCTCCACAGAAAGTCTTTCCCATCTGAGAGAGATTCTTAAAGTGGC 422 db

383 AACAACTCATAGATTGGAGCTGGGTCCAGAAATTTGGGCATACCTGTTTTGGCAATG 442 ov

426 AACRAACTCATAGATTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTTGGCAATG 485

Qy

443 CAGCACTGCAGTGTAAACCTACACACCCTTCGGTTATG
|||||
486 CAGCACTGCAGTGTAAACCTACACACCCTTGCCAATTACATGCTATCACAAGAAC 545

Dd

503 ACTCCAAACATGTTATGCAGAAGGCTTTGTATGATGTGTA CAATGCAGGCACATATTA 562

546 ACTCCAAACATGTTATGCAGAAGGCTTTGTATGATGTGTA CAATGCAGGCACATATTA 605

563 CCCAGGCACTCAGTAATCCTCGGGACGTTATTAAACCAATGTTTTCATCAATGATGATGC 622

db	506	CCCAGGCAC	TCACTAATC	TCTGGGACG	TTATTAAAC	CAATGTTTG	TCATCAAG	AGAGAGC	300
QY	623	GCGGTATAG	CTAGGCAC	CTCCGTTTT	TGGAAAC	CAAGAGATG	CCCATGAAT	TTCCTTCAAT	682

Db
666 GGGGTATAGCTAGGCACITTCGGTTTGGAAACCAAGAAGATGCCCATGAATTCCTCAAT 7255
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

QY

c	78	96	4.0	602	14	CB217660	CB217660 NISC_nb04
c	79	91	3.8	275	13	AA248763	AA248763 j4452.seq
c	80	91	3.8	941	13	BX366515	BX366515 BX366515
c	81	88	3.6	408	28	AQ181035	AQ181035 HS 3219 BA
c	82	68	2.8	711	14	CF178635	CF178635 808125 MA
c	83	66	2.7	75	14	CA513635	CA513635 10-3 1A1
c	84	66	2.7	75	14	CA513637	CA513637 10-3 1A6
c	85	64	2.7	453	28	AQ825501	AQ825501 HS 5499 A
c	86	60	2.5	468	28	AQ603902	AQ603902 HS 2135 A
c	87	59	2.4	705	14	CF178303	CF178303 807357 MA
c	88	56	2.3	667	14	CB422965	CB422965 596130 MA
c	89	56	2.3	687	14	CB428011	CB428011 603459 MA
c	90	53	2.2	415	10	BE079835	BE079835 KC6-BT062
c	91	47	1.9	409	10	BF561189	BF561189 UI-R-C0-h
c	92	47	1.9	550	29	CF5484116	CF5484116 tigr-gss-
c	93	46	1.9	537	10	BF523269	BF523269 UI-R-C3-t
c	94	46	1.9	685	10	BE651621	BE651621 UI-M-BH2
c	95	41	1.7	502	10	BF461120	BF461120 UI-M-CGDP
c	96	41	1.7	532	10	BF404275	BF404275 UI-R-CAL-
c	97	41	1.7	588	10	BF079818	BF079818 H3045D11-
c	98	41	1.7	599	13	C79892	C79892 C79892 Mous
c	99	41	1.7	620	12	BM940384	BM940384 UI-M-CGDP
c	100	41	1.7	644	13	BUE10706	BUE10706 UI-M-FCO-
c	101	41	1.7	672	14	CD350154	CD350154 UI-M-FYO-
c	102	41	1.7	703	13	BQ770634	BQ770634 UI-M-FIO-
c	103	41	1.7	729	14	CF532651	CF532651 UI-M-GHO-
c	104	41	1.7	735	14	CB521515	CB521515 UI-M-GHO-
c	105	41	1.7	771	10	BG066566	BG066566 H3045D11-
c	106	41	1.7	773	14	CB519119	CB519119 UI-M-GHO-
c	107	41	1.7	799	14	CF535655	CF535655 UI-M-GHO-
c	108	41	1.7	1485	11	AK010801	AK010801 Mus muscu
c	109	40	1.7	662	29	AG116029	AG116029 Pan trogl
c	110	38	1.6	318	13	BX129297	BX129297 BX129297
c	111	38	1.6	347	13	BY010203	BY010203 BY010203
c	112	38	1.6	348	13	BY010350	BY010350 BY010350
c	113	38	1.6	391	13	BY050048	BY050048 BY050048
c	114	38	1.6	397	13	BY300193	BY300193 BY300193
c	115	38	1.6	402	13	BY013984	BY013984 BY013984
c	116	38	1.6	412	13	BY018798	BY018798 BY018798
c	117	38	1.5	668	13	BY734416	BY734416 BY734416
c	118	37	1.5	417	10	AW514181	AW514181 hd24906.x
c	119	37	1.5	446	10	BF400191	BF400191 UI-R-CAL-
c	120	34	1.4	750	12	BE1108038	BE1108038 602902424
c	121	32	1.3	618	29	CE119459	CE119459 tigr-gss-
c	122	32	1.3	724	28	BZ247062	BZ247062 CH230-419
c	123	31	1.3	438	14	CB790275	CB790275 AMGNNUC:S
c	124	30	1.2	1977	16	CD013953	CD013953 90137783
c	125	30	1.2	1977	16	AA988223	AA988223 MXRA8_Hum

ALIGNMENTS

[illegible]

```

Db      726  ACACCTGTTGATGTCAGAGGATGCTTGAATGGCAACAATAATTAGACAGACA 785
QY      743  CCAGGCGCCACCTCTGTTTGTTCAGATATTTGGAGATACCTAAGATCTAGAGTCAAA 802
Db      786  CCAGGCGCCACCTCTGTTTGTTCAGATATTTGGAGATACCTAAGATCTAGAGTCAAA 845
QY      803  GTTTAAATGCAAGGGCGTTTCAGATACCTTTGATCCATATCTTGATATATACATTGAGA 862
Db      846  GTTTAAATGCAAGGGCGTTTCAGATACCTTTGATCCATATCTTGATATATACATTGAGA 905
QY      863  TAAAGCTGCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTGAAGCGGAACAGCTTG 922
Db      906  TAAAGCTGCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTGAAGCGGAACAGCTTG 965
QY      923  ATGGAAGAAACTGTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTGAAGCGGAACAG 982
Db      966  ATGGAAGAAACTGTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTGAAGCGGAACAG 1025

RESULT 2
LOCUS   BX354198                                1006 bp    mRNA    linear    EST 05-MAY-2003
DEFINITION   BX354198 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
ACCESSION   BX354198
VERSION     BX354198.1  GI:30367814
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES             source
1..1006
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODC016Y21"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/notes="First strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match      34.4%; Score 830; DB 13; Length 1006;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 880; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      30  CGAGGGGATGAGCGAGCGCGCGGTCAGAGTTGAACATACCATAGTTGCAAA 89
Db      67  CGAGGGGATGAGCGAGCGCGCGGTCAGAGTTGAACATACCATAGTTGCAAA 126
QY      90  AGCTTCTGAATCTTCAGACCATCAGCTATCAGAAATCAGCTCGCAGCTCCGAGG 149
Db      127 AGCTTCTGAATCTTCAGACCATCAGCTATCAGAAATCAGCTCGCAGCTCCGAGG 186

```

```

QY      150  CTACCTCGGAGACATGAGTACAGGTTCTGCAGCTGGGGTCTGTCTTCATTGAATGA 209
Db      187  CTCACCTGGAGACATGAGTACAGGTTCTGCAGCTGGGGTCTGTCTTCATTGAATGA 246
QY      210  TGTGTCAAAATCACACACTTCTTTTAGGACAGTACCTGTGTGTAGTTTATTGAGTTC 269
Db      247  TGTGTCAAAATCACACACTTCTTTTAGGACAGTACCTGTGTGTAGTTTATTGAGTTC 306
QY      270  ATCTGTACCTGATATAATCAAAACCATCACACAAAAGGATCAAGCCCTAGTGTATG 329
Db      307  ATCTGTACCTGATATAATCAAAACCATCACACAAAAGGATCAAGCCCTAGTGTATG 366
QY      330  CGCTCTCTCCACAGAAAGTTCTTTTCCCATCTCGAAGAAGATTTGTCTTAAGTGGCAACAAAC 389
Db      367  CGCTCTCTCCACAGAAAGTTCTTTTCCCATCTCGAAGAAGATTTGTCTTAAGTGGCAACAAAC 426
QY      390  TCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGCAATACCTGTTTGGCAATGACGACT 449
Db      427  TCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGCAATACCTGTTTGGCAATGACGACT 486
QY      450  GCAGTGTTTAACTACACACACCTCTTCCCAATTTACATGCTATCATGACATGACACTCCAA 509
Db      487  GCAGTGTTTAACTACACACACCTCTTCCCAATTTACATGCTATCATGACATGACACTCCAA 546
QY      510  AACATGTCATGAGAGGCTTTTGTATGATGTGTACAATGCAAGCACATATTACCCAGGC 569
Db      547  AACATGTCATGAGAGGCTTTTGTATGATGTGTACAATGCAAGCACATATTACCCAGGC 606
QY      570  ACTCAGTAACTCTGGGAGCTTATTAAACCAATGTTTGTATCAATAGATGCGGCGTAT 629
Db      607  ACTCAGTAACTCTGGGAGCTTATTAAACCAATGTTTGTATCAATAGATGCGGCGTAT 666
QY      630  AGCTAGGACCTCCGTTTGGAAACCAAGAGATGCCATGAATTCCTTCAATACACTGT 689
Db      667  AGCTAGGACCTCCGTTTGGAAACCAAGAGATGCCATGAATTCCTTCAATACACTGT 726
QY      690  TGATGCTATGCAAGAGCATGCTTGAATGGCAGCAATAAATTAGACACACACCCAGGC 749
Db      727  TGATGCTATGCAAGAGCATGCTTGAATGGCAGCAATAAATTAGACACACACCCAGGC 786
QY      750  CACCACTCTGTTGTGATATTTGGAGGATACCTAAGATCTAGAGTCAAAATGTTTAA 809
Db      787  CACCACTCTGTTGTGATATTTGGAGGATACCTAAGATCTAGAGTCAAAATGTTTAA 846
QY      810  TTGCAAGGGCGTTTCAGATACCTTTGATCCATATCTTGTATATAACATGGAGATAAGGC 869
Db      847  TTGCAAGGGCGTTTCAGATACCTTTGATCCATATCTTGTATATAACATGGAGATAAGGC 906
QY      870  TGCTCAGAGTGTCAACAGGCAATTTGGAGCAGTTTGTGAAGC 910
Db      907  TGCTCAGAGTGTCAACAGGCAATTTGGAGCAGTTTGTGAAGC 947

```

```

RESULT 3
BX418054
LOCUS       BX418054 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION   CSODF001YG23 5-PRIME, mRNA sequence.
ACCESSION   BX418054
VERSION     BX418054.1  GI:30646401
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 952)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France

```

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 4596.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DF001AD12QPI&cluster=4596.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DF001AD12QPI.
Location/Qualifiers
1..952
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="CS0DF001YG23"
/tissue.type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

FEATURES

source
Query Match 32.5%; Score 785; DB 13; Length 952;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 885; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 23 CGCGCGGAGGGGATGGAGCGAGCGCGGGTCAAGTTCGAACATGACCATAG 82
DB 66 CGCGCGGAGGGGATGGAGCGAGCGCGGGTCAAGTTCGAACATGACCATAG 125
QY 83 TTGACAAAGCTTCTGAATCTTCAGACCATCAGCCATCAGATTCAGATTCGCG 142
DB 126 TTGACAAAGCTTCTGAATCTTCAGACCATCAGCCATCAGATTCAGATTCGCG 185
QY 143 AGGCAGTCTCACCTGGAGACATGATGACGAGTTCGCGAGCTGGGGTGTCTTCAT 202
DB 186 AGGCAGTCTCACCTGGAGACATGATGACGAGTTCGCGAGCTGGGGTGTCTTCAT 245
QY 203 TGAATGATGTCAAAATCAGACATCTTCTTTAGGACCACTGCTGTGTAGTTTATT 262
DB 246 TGAATGATGTCAAAATCAGACATCTTCTTTAGGACCACTGCTGTGTAGTTTATT 305
QY 263 CGAGTTCATCTGTACCTGATAAATCAAAACCATCACCACAAAGGATCAAGCCCTAGGTG 322
DB 306 CGAGTTCATCTGTACCTGATAAATCAAAACCATCACCACAAAGGATCAAGCCCTAGGTG 365
QY 323 ATGGCATCGCTCTCCACAGAAAGTCTTTTCCCATCTGAGAAAGTTCCTTAAGTGGC 382
DB 366 ATGGCATCGCTCTCCACAGAAAGTCTTTTCCCATCTGAGAAAGTTCCTTAAGTGGC 425
QY 383 AACAACTCATAGTGTGGAGCTGGCTCCAGAAATTTGGGCAATACCTGTTTGGCAATG 442
DB 426 AACAACTCATAGTGTGGAGCTGGCTCCAGAAATTTGGGCAATACCTGTTTGGCAATG 485
QY 443 CAGCACTGCAGTGTAACTACACACACCATCTTTCGCAATATACATGCTATCAATGAAC 502
DB 486 CAGCACTGCAGTGTAACTACACACACCATCTTTCGCAATATACATGCTATCAATGAAC 545
QY 503 ACTCCAAACATGTCATGAGAGGCTTTTGTATGATGTGTACATGCAAGCACATATTA 562
DB 546 ACTCCAAACATGTCATGAGAGGCTTTTGTATGATGTGTACATGCAAGCACATATTA 605
QY 563 CCAGGCACTCAGTAATCTCGGGAGCTTATTAACCAATGTTTGTATCAATGAGATGC 622
DB 606 CCAGGCACTCAGTAATCTCGGGAGCTTATTAACCAATGTTTGTATCAATGAGATGC 665
QY 623 GCGGTATAGTAGGACCTTCGTTTGGAAACCAAGAGATGCCCATGAAATCTTCAAT 682
DB 666 GCGGTATAGTAGGACCTTCGTTTGGAAACCAAGAGATGCCCATGAAATCTTCAAT 725

ORIGIN

Query Match 31.3%; Score 756; DB 12; Length 1063;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 806; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 513 ATGTCATCAGAAGGCTTTTGTATGATGTGTACATGCAAGCACATATACCAGGCACT 572
DB 17 ATGTCATCAGAAGGCTTTTGTATGATGTGTACATGCAAGCACATATACCAGGCACT 76
QY 573 CAGTAATCTCGGGAGCTTATTAACCAATGTTTGTATCAATGAGATGCGGCGTATAGC 632

QY 683 ACACGTGTTGATGCTATGCAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACA 742
DB 726 ACACGTGTTGATGCTATGCAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACA 785
QY 743 CCCAGGCCACCACTCTGTTTGTCTAGATATTTGGAGGATACCTAAGATCTAGAGTCAAAAT 802
DB 786 CCCAGGCCACCACTCTGTTTGTCTAGATATTTGGAGGATACCTAAGATCTAGAGTCAAAAT 845
QY 803 GTTTAAATGCAAGGCGCTTTCAGATACCTTTTGATCCATATCTTGATATAAATTGGAGA 862
DB 846 GTTTAAATGCAAGGCGCTTTCAGATACCTTTTGATCCATATCTTGATATAAATTGGAGA 905
QY 863 TAAAGGCTGCTCAGAGTGTCAACAAGGCATTCGAGCAGTTTGTGAAG 909
DB 906 TAAAGGCTGCTCAGAGTGTCAACAAGGCATTCGAGCAGTTTGTGAAG 952

RESULT 4

BM920091 1063 bp mRNA linear EST 12-MAR-2002
AGENCOURT 6706520 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5749963
5', mRNA sequence.
ACCESSION BM920091
VERSION BM920091.1 GI:19370470
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1063)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1AM12780 row: c column: 20
High quality sequence stop: 760.

FEATURES

source
1..1063
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="IMAGE:5749963"
/lab_host="DH10B"
/clone_lib="NIH_MGC_120"
/note="Organ: pooled pancreas and spleen; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of spleen and pancreas from 28 yo
male. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 1.5 kb, insert size range 1-2.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 025. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 31.3%; Score 756; DB 12; Length 1063;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 806; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 513 ATGTCATCAGAAGGCTTTTGTATGATGTGTACATGCAAGCACATATACCAGGCACT 572
DB 17 ATGTCATCAGAAGGCTTTTGTATGATGTGTACATGCAAGCACATATACCAGGCACT 76
QY 573 CAGTAATCTCGGGAGCTTATTAACCAATGTTTGTATCAATGAGATGCGGCGTATAGC 632

Db 77 CAGTAATCTCGGAGCGTTAATAAACCAATGTTTTCATCAATGAGATGGCGGTATAGC 136
QY 633 TAGGCACCTCCGTTTGGAAACCAAGAGATGCCATGAATTCCTTCAATACACTGTTGA 692
Db 137 TAGGCACCTCCGTTTGGAAACCAAGAGATGCCATGAATTCCTTCAATACACTGTTGA 196
QY 693 TGCTATGCAAGAACGATGCTTGAATGGCAGCAATAAATAGACAGACACACCCAGGCCAC 752
Db 197 TGCTATGCAAGAACGATGCTTGAATGGCAGCAATAAATAGACAGACACACCCAGGCCAC 256
QY 753 CACTCTTGTGTCAGATATTTGGAGATACCTAAGATCTAGAGTCAAAATGTTTAAATTG 812
Db 257 CACTCTTGTGTCAGATATTTGGAGATACCTAAGATCTAGAGTCAAAATGTTTAAATTG 316
QY 813 CAAGGCGGTTTCAGATPACTTTGATCCATATCTTGATATAAATTCGAGATAAAGCTGC 872
Db 317 CAAGGCGGTTTCAGATPACTTTGATCCATATCTTGATATAAATTCGAGATAAAGCTGC 376
QY 873 TCAGAGTGTCAACAAAGGCAATGGAGCAGTTTGTGAAGCCGGAACAGCTTGATGGAGAAA 932
Db 377 TCAGAGTGTCAACAAAGGCAATGGAGCAGTTTGTGAAGCCGGAACAGCTTGATGGAGAAA 436
QY 933 CTCGTACAGTGCAGCAAGTGTAAAGATGTTCCAGCTTCAAAGAGGTTCACTATCCA 992
Db 437 CTCGTACAGTGCAGCAAGTGTAAAGATGTTCCAGCTTCAAAGAGGTTCACTATCCA 496
QY 993 TAGATCCTCTAATGTTTACACTTCTCTGAAACGTTTGTGCAAAATTTTACCGTGGAAA 1052
Db 497 TAGATCCTCTAATGTTTACACTTCTCTGAAACGTTTGTGCAAAATTTTACCGTGGAAA 556
QY 1053 AATTGCTAAGGATGTGAATACCTCGAGTATCTTGATATTCGGCCATATATGCTCAACC 1112
Db 557 AATTGCTAAGGATGTGAATACCTCGAGTATCTTGATATTCGGCCATATATGCTCAACC 616
QY 1113 CAACGAGAGCAATGTCTAGCTCTGATGACAGTGTGTCACACTGTTTAAATTG 1172
Db 617 CAACGAGAGCAATGTCTAGCTCTGATGACAGTGTGTCACACTGTTTAAATTG 676
QY 1173 CCATGCTGCCATTAATCTCTGATCAATAAAGTACGATGCGCTTGGTATCAATGAA 1232
Db 677 CCATGCTGCCATTAATCTCTGATCAATAAAGTACGATGCGCTTGGTATCAATGAA 736
QY 1233 TGACTCCATGATCTACAGTGAATATGATGCGGTACTCAGCCAAACAGCCTATGTCCT 1292
Db 737 TGACTCCATGATCTACAGTGAATATGATGCGGTACTCAGCCAAACAGCCTATGTCCT 796
QY 1293 CTTTATATCAGGTCCTATGATGAA 1319
Db 797 CTTTATATCAGGTCCTATGATGAA 823

RESULT 5
LOCUS BUI73371 894 bp mRNA linear EST 04-SEP-2002
DEFINITION AGNCOURT_7984636 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6174750
5', mRNA sequence.
ACCESSION BUI73371
VERSION BUI73371.1 GI:22687355
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 894)
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM13548 row: g column: 07
High quality sequence stop: 670.
Location/Qualifiers
1. .894
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6174750"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 71"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: Sall; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.1 kb."

ORIGIN

Query Match 30.8%; Score 744; DB 13; Length 894;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 794; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 109 CCATCAGCCTATCAGATCAGCTCGGAGTCTCAGCTCGGAGATGAT 168
Db 1 CCATCAGCCTATCAGATCAGCTCGGAGTCTCAGCTCGGAGATGAT 60
QY 169 GCAGGTTCTGCCAGCTGGGGTCTGCTTCATTAATGAATGATGTCAAATCACACATT 228
Db 61 GCAGGTTCTGCCAGCTGGGGTCTGCTTCATTAATGAATGATGTCAAATCACACATT 120
QY 229 TCTTTAGACCACTACCTGGTCTGTAGTTTATTCGAGTTCTATCTGACCTGATATAATCA 288
Db 121 TCTTTAGACCACTACCTGGTCTGTAGTTTATTCGAGTTCTATCTGACCTGATATAATCA 180
QY 289 AAACCATCACCAAAAGATCAAGCCTTAGGTGATGCGCTCCTCCACAGAAAGTT 348
Db 181 AAACCATCACCAAAAGATCAAGCCTTAGGTGATGCGCTCCTCCACAGAAAGTT 240
QY 349 CTTTTCCTCATCTGAGAAGATTGCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGG 408
Db 241 CTTTTCCTCATCTGAGAAGATTGCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGG 300
QY 409 CTCAGAAATTTGGGCAATACCTGTTTCCCAATGAGCACTGCACTGTTTAACTACACA 468
Db 301 CTCAGAAATTTGGGCAATACCTGTTTCCCAATGAGCACTGCACTGTTTAACTACACA 360
QY 469 CCACCTCTTGGCAATACCTGTTTCCCAATGAGCACTGCACTGTTTAACTACACA 528
Db 361 CCACCTCTTGGCAATACCTGTTTCCCAATGAGCACTGCACTGTTTAACTACACA 420
QY 529 TTTTGTATGATGTATCAATGCAAGCACATATTACCCAGGCACTCAGTAATCTCGGGAC 588
Db 421 TTTTGTATGATGTATCAATGCAAGCACATATTACCCAGGCACTCAGTAATCTCGGGAC 480
QY 589 GTTATTAACCAATGTTTGTTCATCAATGAGATGGGGGTATAGCTAGGCACTCCGTTTT 648
Db 481 GTTATTAACCAATGTTTGTTCATCAATGAGATGGGGGTATAGCTAGGCACTCCGTTTT 540
QY 649 GGAAACCAAGAGATGCCATGAATTCCTCAATACACTGTTGATGCTATGCAAGAGCA 708
Db 541 GGAAACCAAGAGATGCCATGAATTCCTCAATACACTGTTGATGCTATGCAAGAGCA 600
QY 709 TGCCTGAATGGCAGCAATAAATAGACAGACACACCCAGGCCACCTCTGTTGTCAG 768
Db 601 TGCCTGAATGGCAGCAATAAATAGACAGACACACCCAGGCCACCTCTGTTGTCAG 660
QY 769 ATATTGGAGATACCTAAGATCTAGAGTCAAAATGTTTAAATTCGAAGGCGTTTCAGAT 828
Db 661 ATATTGGAGATACCTAAGATCTAGAGTCAAAATGTTTAAATTCGAAGGCGTTTCAGAT 720
QY 829 ACTTTTGCATATCTTGTATATAAATTCGAGATAAAGGCTGCTCAGAGTGTCAACAAG 888

Db	721	ACTTTGATCCATATCTTGATATATAACATTGGAGATAAAGGCTGCTCAGAGTGTCAACAAG	780
Qy	889	GCATTGGAGCAGTTT	903
Db	781	GCATTGGAGCAGTTT	795
RESULT 6			
BUI95574			
LOCUS	BUI95574	880 bp	linear mRNA
DEFINITION	AGENCOURT_7967650 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6171493		EST 04-SEP-2002
ACCESSION	BUI95574		5', mRNA sequence.
VERSION	BUI95574.1	GI:22709558	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 880)		
TITLE	NIH-MGC http://mgi.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-i@mail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Life Technologies, Inc.		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov		
	Plate: L1AMI3539 row: o column: 14		
	High quality sequence stop: 710.		
FEATURES	Location/Qualifiers		
source	1..880		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6171493"		
	/tissue_type="retinoblastoma"		
	/lab_host="DH10B (phage-resistant)"		
	/clone_lib="NIH_MGC_67"		
	/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies."		
ORIGIN			
Query Match	30.5%	Score 735;	DB 13; Length 880;
Best Local Similarity	99.9%	Pred. No. 0;	
Matches	785;	Conservative 0;	Mismatches 1; Indels 0; Gaps 0;
Qy	42	AGCGAGCGCCGAGCGGCTGAGAGTTGAACAATGACCATAGTTCACAAAGCTTCTGAATC	101
Db	1	AGCGAGCGCCGAGCGGCTGAGAGTTGAACAATGACCATAGTTCACAAAGCTTCTGAATC	60
Qy	102	TTTCAGACCCATCAGCCTATCAGATACAGCCCTGGCAGTCCGAGGAGTCTCACCTGGAGA	161
Db	61	TTTCAGACCCATCAGCCTATCAGATACAGCCCTGGCAGTCCGAGGAGTCTCACCTGGAGA	120
Qy	162	CATGATGACAGTTCTGCAGTGGGGCTGCTCTTCATTGATGATGTCAGAAATCA	221
Db	121	CATGATGACAGTTCTGCAGTGGGGCTGCTCTTCATTGATGATGTCAGAAATCA	180
Qy	222	CACACTTTCTTTAGACACAGTACCTGGTGTCTGTTATTTTCGAGTTCATCTGTACCTGA	281
Db	181	CACACTTTCTTTAGACACAGTACCTGGTGTCTGTTATTTTCGAGTTCATCTGTACCTGA	240
Qy	282	TAAATCAAAACCATCAACCAAAAGAGATCAAGCCCTAGTGTGATGTCATCGCTCCACCA	341
Db	241	TAAATCAAAACCATCAACCAAAAGAGATCAAGCCCTAGTGTGATGTCATCGCTCCACCA	300
Qy	342	GAAAGTCTTTTCCCATCTGAGAGATTTGTCTTAAGTGGCAACAACTCATAGAGTTGG	401
Db	301	GAAAGTCTTTTCCCATCTGAGAGATTTGTCTTAAGTGGCAACAACTCATAGAGTTGG	360
Qy	402	AGCTGGGCTCCAGAA"TTGGGCAATACCTGTTTGGCAATGAGCACTGAGTGTGTTAAC	461
Db	361	AGCTGGGCTCCAGAA"TTGGGCAATACCTGTTTGGCAATGAGCACTGAGTGTGTTAAC	420
Qy	462	CTACACACACCTCTTGGCAATTACATGCTATCAGATGAACACTCCAAAACATGTCTATGC	521
Db	421	CTACACACACCTCTTGGCAATTACATGCTATCAGATGAACACTCCAAAACATGTCTATGC	480
Qy	522	AGAAGGCTTTTGTATGATGTGTACAAATGCAAGCACATATTAACCCAGGCACTCAGTAATCC	581
Db	481	AGAAGGCTTTTGTATGATGTGTACAAATGCAAGCACATATTAACCCAGGCACTCAGTAATCC	540
Qy	582	TGGGAGCGTTTATAACCAATGTTTGTATCAATGAGATGCGGCGTATAGTAGGACCT	641
Db	541	TGGGAGCGTTTATAACCAATGTTTGTATCAATGAGATGCGGCGTATAGTAGGACCT	600
Qy	642	CGTTTTGGAAACCAAGAGATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGCA	701
Db	601	CGTTTTGGAAACCAAGAGATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGCA	660
Qy	702	GAAAGCATGCTTGAATGGCAGCAATTAATTTAGACAGACACACCCAGGCCACCTCTTGT	761
Db	661	GAAAGCATGCTTGAATGGCAGCAATTAATTTAGACAGACACACCCAGGCCACCTCTTGT	720
Qy	762	TTTTCAGATATTTGGAGGATACCTAGATCTAGATCAAAATGTTTAAATTCGAAGGCGCT	821
Db	721	TTTTCAGATATTTGGAGGATACCTAGATCTAGATCAAAATGTTTAAATTCGAAGGCGCT	780
Qy	822	TTTCAGA 827	
Db	781	TTTCAGA 786	
RESULT 7			
BX451363			
LOCUS	BX451363	918 bp	mRNA linear
DEFINITION	BX451363 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone		EST 22-MAY-2003
ACCESSION	CU000282F10 5-PRIME, mRNA sequence.		
VERSION	BX451363		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 918)		
TITLE	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.		
JOURNAL	Full-length cDNA libraries and normalization		
COMMENT	Unpublished (2001)		
	Contact: Genoscope		
	Genoscope - Centre National de Sequencage		
	BP 191 91006 EVRY cedex - France		
	Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr		
	Library was constructed by Life Technologies, a division of		
	Invitrogen. This sequence belongs to sequence cluster 5254.r For		
	more information about this cluster, see		
	http://www.genoscope.cns.fr/		
	cgf-hp1/cluster.cgi?seq=CS0BAF001ZF08_AF00063_1&cluster=5254.r.		
	Contact : Feng Liang Email : fliang@lifetech.com URL :		
	http://fulllength.invitrogen.com/ Invitrogen Corporation 1600		
	Paradise Avenue Genoscope sequence ID : CS0BAF001ZF08_AF00063_1.		
FEATURES	Location/Qualifiers		
source	1..918		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CU000282F10"		
	/tissue_type="NEUROBLASTOMA"		
	/clone_lib="Homo sapiens NEUROBLASTOMA"		

/note="vector: pcwvSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pcwvSPORT 6 vector. Library was not normalized."

ORIGIN

```

Query Match      29.0%; Score 699; DB 13; Length 918;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 749; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 864 AAAGCGTGTCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTCAAGCCGGAACAGCTTGA 923
Db 21 AAAGCGTGTCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTCAAGCCGGAACAGCTTGA 80
QY 924 TGGAGAAAACCTCGTACAGTGCAGCAGTGTAAAGAGATGGTTCAGCTTCAAGAGGTT 983
Db 81 TGGAGAAAACCTCGTACAGTGCAGCAGTGTAAAGAGATGGTTCAGCTTCAAGAGGTT 140
QY 984 CACTATCCATAGATCTCTAAATGTTCTTACACATTTCTCTGAAACGTTTGTGCAATTTTAC 1043
Db 141 CACTATCCATAGATCTCTAAATGTTCTTACACATTTCTCTGAAACGTTTGTGCAATTTTAC 200
QY 1044 CGGTGAAAAATGCTAAGGATGTGAATACCTGTAGTATCTTGTATTTTCGGCCATATAT 1103
Db 201 CGGTGAAAAATGCTAAGGATGTGAATACCTGTAGTATCTTGTATTTTCGGCCATATAT 260
QY 1104 GTCTCAACCCAGGAGAGCCAAATGTCTACGCTCTGTATGTCAGTCTGGTCCACACTGG 1163
Db 261 GTCTCAACCCAGGAGAGCCAAATGTCTACGCTCTGTATGTCAGTCTGGTCCACACTGG 320
QY 1164 TTTTAATTGCCATGCTGGCCATTAATCTCTGTACATAAAAGCTAGCAATGGCTCTGGTA 1223
Db 321 TTTTAATTGCCATGCTGGCCATTAATCTCTGTACATAAAAGCTAGCAATGGCTCTGGTA 380
QY 1224 TCAATGATGACTCCATTTGATCTACAGTGATATTAGATCGTACTCAGCCAAACAGC 1283
Db 381 TCAATGATGACTCCATTTGATCTACAGTGATATTAGATCGTACTCAGCCAAACAGC 440
QY 1284 CTATGTGCTCTTTTATATACAGTCCCATGATGTGAAATGAGAGTGAACTTACTCATCC 1343
Db 441 CTATGTGCTCTTTTATATACAGTCCCATGATGTGAAATGAGAGTGAACTTACTCATCC 500
QY 1344 CACCATAGCCCGGCCAGTCTCTCCCGCCCGCTCATCATCAGTCAAGCGGTTGTCAACAA 1403
Db 501 CACCATAGCCCGGCCAGTCTCTCCCGCCCGCTCATCATCAGTCAAGCGGTTGTCAACAA 560
QY 1404 CAACAGGCTGGCCAGGCTTTATCGGACCAAGCTTCCCTCTCATATGATAAAGAAATCC 1463
Db 561 CAACAGGCTGGCCAGGCTTTATCGGACCAAGCTTCCCTCTCATATGATAAAGAAATCC 620
QY 1464 ACCTCACTTAAATGGAGCTGGACCATTTGAAGACACGCAAGCAGTTCATGTGAGTCC 1523
Db 621 ACCTCACTTAAATGGAGCTGGACCATTTGAAGACACGCAAGCAGTTCATGTGAGTCC 680
QY 1524 TAACGGAAATCCAGTGTCAACAGGGGTAGTCCCTGTTAATGCTTCAGTCTTGTCCTCAAA 1583
Db 681 TAACGGAAATCCAGTGTCAACAGGGGTAGTCCCTGTTAATGCTTCAGTCTTGTCCTCAAA 740
QY 1584 CTGTCAGTTAATAGGCTCTCAGTGATGCC 1613
Db 741 CTGTCAGTTAATAGGCTCTCAGTGATGCC 770

```

RESULT 8
 AUI24746
 LOCUS AUI24746 NT2RM4 Homo sapiens cDNA clone NT2RM4000488 5', mRNA
 DEFINITION AUI24746 851 bp mRNA linear EST 01-AUG-2002
 sequence.
 ACCESSION AUI24746
 VERSION AUI24746.1 GI:10949462
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 851)

REFERENCE
AUTHORS

Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J.,
 Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S. and
 Isogai, T.
 HRI human cDNA project (Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S.,
 Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T.,
 Suzuki, Y., Sugano, S., Isogai, T.)
 Unpublished (2000)

JOURNAL

Contact: Takao Isogai

COMMENT

Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3975
 Fax: 81-438-52-3986
 Email: genomics@hri.co.jp

HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.

FEATURES

Location/Qualifiers

```

1..851
Source
1..851
Location/Qualifiers
1..851
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RM4000488"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/clone_lib="NT2RM4"
/note="Vector: pME18SPL3; mRNA from uninduced NT2 neuronal
precursor cells"

```

ORIGIN

```

Query Match      28.7%; Score 692; DB 9; Length 851;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 692; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1339 CATCCCAACCCATAGCCCGCCAGTCTCTCCCGCCCGTCATCAGTCAGCGGTTGTC 1398
Db 2 CATCCCAACCCATAGCCCGCCAGTCTCTCCCGCCCGTCATCAGTCAGCGGTTGTC 61
QY 1399 ACCAACAACAGGCTGCGCCAGGCTTTATCGACCAACAGCTTCCCTCTCACATGATAAG 1458
Db 62 ACCAACAACAGGCTGCGCCAGGCTTTATCGACCAACAGCTTCCCTCTCACATGATAAG 121
QY 1459 AATCCACTCTATTAAATGGGACTGGACCAATGAAAGACACGCCAAGCAGTTCCATGTCG 1518
Db 122 AATCCACTCTATTAAATGGGACTGGACCAATGAAAGACACGCCAAGCAGTTCCATGTCG 181
QY 1519 AGTCCTAACGGGAATTCAGTGTCAACAGGCTAGTCTGTTAATGCTTCAAGTCTGTC 1578
Db 182 AGTCCTAACGGGAATTCAGTGTCAACAGGCTAGTCTGTTAATGCTTCAAGTCTGTC 241
QY 1579 CAAATCTGTCAGTAAATAGTCTCTCAGTGTATCCAGAAACATCCCTAAGAAACAAAAATT 1638
Db 242 CAAATCTGTCAGTAAATAGTCTCTCAGTGTATCCAGAAACATCCCTAAGAAACAAAAATT 301
QY 1639 ACATCATGATTTACACAAGTTGCCCTGTTGCGCAGTGTGAGTCTCAACCTTCAAT 1698
Db 302 ACATCATGATTTACACAAGTTGCCCTGTTGCGCAGTGTGAGTCTCAACCTTCAAT 361
QY 1699 AGTAATTTCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCATTCATTCGCA 1758
Db 362 AGTAATTTCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCATTCATTCGCA 421
QY 1759 GTACAGTCTACCTCGAACGCAATCTACGATGTCTAGTCTTCTAGTAAAGTAAACAAACCCATC 1818
Db 422 GTACAGTCTACCTCGAACGCAATCTACGATGTCTAGTCTTCTAGTAAAGTAAACAAACCCATC 481
QY 1819 CCGCGAGTGAATCTGCTCCCGCCCGGTGATGAATGGCAATCCAAAGCTGAATCCAGC 1878

```

Db	482	CCCGCAGTGAATCTCTCTCCAGCCCGTGATGAATGGCAAAATCCAAAGTGAACCTCCAGC	541
Qy	1879	GTGCTGGTGCCTATGGCGCGAGTCTCTGAGGACTCTGACGAGAGTCAAAGGGGCTG	1938
Db	542	GTGCTGGTGCCTATGGCGCGAGTCTCTGAGGACTCTGACGAGAGTCAAAGGGGCTG	601
Qy	1939	GGCAAGGAGAAATGGGATTGGTAGCTTGAGTCCCACTCTCCGGCCCAAGATGCCGA	1998
Db	602	GGCAAGGAGAAATGGGATTGGTAGCTTGAGTCCCACTCTCCGGCCCAAGATGCCGA	661
Qy	1999	GATGAGAGGCCACTCCGCAAGAGCTTCAAGA	2030
Db	662	GATGAGAGGCCACTCCGCAAGAGCTTCAAGA	693

RESULT 9					
AU130866					
LOCUS	AU130866	763 bp	mRNA	linear	EST_01-AUG-2002
DEFINITION	AU130866 NT2RP3 Homo sapiens cDNA clone NT2RF3001577 5', mRNA sequence.				

ACCESSION AUI30866
 VERSION AUI30866.1 GI:10991220
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 763)
 Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,
 Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, I., Sugano, S. and
 Isozaki, T.

TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

```

FEATURES
source
Location/Qualifiers
1..763
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP3001577"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/clone_lib="NT2RP3"
/notes="vector: PME18SPL3; mRNA from NT2 neuronal precursor
cells; beta-actin; acid (pA) induction"

```

CELLS AFTER 2-WEEKS RETROVIRAL SELECT (AUG. 1982)									
ORIGIN									
	Query Match	27.9%;	Score 672;	DB 9;	Length 763;				
	Best Local Similarity	100.0%;	Pred. No. 0;						
	Matches 672;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1166	TTAATTGGCCATGCTGGGCCATTACTTCTGCTACATAAAAGCTAGCAATGGCCCTCTGGTATC							1222
ST	1	TTAATTGGCCCTGCTGGGCCATTACTTCTGCTACATAAAAGCTAGCAATGGCCCTCTGGTATC							60

DB	Qy	Qx	Qz	Qy	Qx	Qz
1	AAATGATGACATCCCAATGTTATCTACACAGTGATATTAGATCGGTACTCAGCCAAACAGCT	1286	ATGTGCTCTTTTATATCAGTCCCATGCTGAAAAATCGAGGTGAATTTACTCATCCCA	1341	ATGTGCTCTTTTATATCAGTCCCATGCTGAAAAATCGAGGTGAATTTACTCATCCCA	1801
2	AAATGATGACATCCCAATGTTATCTACACAGTGATATTAGATCGGTACTCAGCCAAACAGCT	1226	ATGTGCTCTTTTATATCAGTCCCATGCTGAAAAATCGAGGTGAATTTACTCATCCCA	1286	ATGTGCTCTTTTATATCAGTCCCATGCTGAAAAATCGAGGTGAATTTACTCATCCCA	1801
3	AAATGATGACATCCCAATGTTATCTACACAGTGATATTAGATCGGTACTCAGCCAAACAGCT	61	AAATGATGACATCCCAATGTTATCTACACAGTGATATTAGATCGGTACTCAGCCAAACAGCT	120	AAATGATGACATCCCAATGTTATCTACACAGTGATATTAGATCGGTACTCAGCCAAACAGCT	120

1346	QY	CCCATAGCCCGGCGCAGTCTCTCCGCGCCCGGTATCATAGTCAGCGGGTGTGCACCAACA	1400
181	DB	CCCATAGCCCGGCGCAGTCTCTCCGCGCCCGGTATCATAGTCAGCGGGTGTGCACCAACA	240
1406	QY	AACAGGCTCGGCAGGCTTTATCGGACCAACAGTTTCCCTCTCATCATGATAAAGAAATCCAC	1465
241	DB	AACAGGCTCGGCAGGCTTTATCGGACCAACAGTTTCCCTCTCATCATGATAAAGAAATCCAC	300
1466	QY	CTCACCTAAATGGGACTGGACCAATTCAAAGACACGCGCAAGCAGTTCATGTGCGAGTCCTA	1525
301	DB	CTCACCTAAATGGGACTGGACCAATTCAAAGACACGCGCAAGCAGTTCATGTGCGAGTCCTA	360
1526	QY	ACGGGAATTCAGTGTCAACAGGGCTAGTCTGTTTAATGCTTCAGCTTCTGTGTCCTCAAACT	1585
361	DB	ACGGGAATTCAGTGTCAACAGGGCTAGTCTGTTTAATGCTTCAGCTTCTGTGTCCTCAAACT	420
1586	QY	GGTCAGTTAATAGGTCTCAGTGCATCCAGAACATCTTAAGAAAACAAAAATTCAAATCA	1645
421	DB	GGTCAGTTAATAGGTCTCAGTGCATCCAGAACATCTTAAGAAAACAAAAATTCAAATCA	480
1646	QY	GTATTACAAACAAGTTGCGCTGTTCGCGAGTGTGAGTCTCAACCTTAACCTTCATAGTAATT	1705
481	DB	GTATTACAAACAAGTTGCGCTGTTCGCGAGTGTGAGTCTCAACCTTAACCTTCATAGTAATT	540
1706	QY	CTTTGGGAACCCCTAACAAGCCGTTCCCTCTTCTACATTACCAATTCGCGAGTACAGT	1765
541	DB	CTTTGGGAACCCCTAACAAGCCGTTCCCTCTTCTACATTACCAATTCGCGAGTACAGT	600
1766	QY	CTACCTCGAACGGCATCTACGATGTCAAGTTTCTAGTAAAGTAACAAAACGGATCCCGCGCA	1825
601	DB	CTACCTCGAACGGCATCTACGATGTCAAGTTTCTAGTAAAGTAACAAAACGGATCCCGCGCA	660
1826	QY	GTGAATCCTGCT	1837
661	DB	GTGAATCCTGCT	672

RESULT 10	AUI24710	822 bp	linear	EST 01-AUG-2002
LOCUS	AUI24710	NT2RM4 Homo sapiens cDNA clone NT2RM4000431 5', mRNA		
DEFINITION	AUI24710	sequence.		
ACCESSION	AUI24710	GI:1C949425		
VERSION	AUI24710.1			
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 822)			
AUTHORS	Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J., Nakamura, K., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S. and Isogai, T.			
TITLE	HRI human cDNA project (Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S., Isogai, T.)			

Unpublished (2000)
Contact: takao isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human CDNA project; 5' - & 3' - end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1. .822
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

```

/clone="NT2RM4000431"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/clone_lib="NT2RM4"
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
precursor cells"

ORIGIN
Query Match      27.5%; Score 663; DB 9; Length 822;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 713; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 856 TTGGAGATAAAGCTGCTCAGAGTGTCAACAAGGCATTGAGCAGATTTGTGAAGCCGGAA 915
Db 1 TTGGAGATAAAGCTGCTCAGAGTGTCAACAAGGCATTGAGCAGATTTGTGAAGCCGGAA 60
QY 916 CAGCTTGATGGAGAAACCTGTCACAGTGCAGCAAGTGTAAAGAATGGTTCAGATTCA 975
Db 61 CAGCTTGATGGAGAAACCTGTCACAGTGCAGCAAGTGTAAAGAATGGTTCAGATTCA 120
QY 976 AAGAGGTTCACTATCCATAGATCCTCTAATGTTCTTACACTTTCTCTGAAACGTTTTCGA 1035
Db 121 AAGAGGTTCACTATCCATAGATCCTCTAATGTTCTTACACTTTCTCTGAAACGTTTTCGA 180
QY 1036 AATTTTACCGTGGAAAAATTCGTAAGGATGCAATACCCCTGAGTATCTTGATATTCGG 1095
Db 181 AATTTTACCGTGGAAAAATTCGTAAGGATGCAATACCCCTGAGTATCTTGATATTCGG 240
QY 1096 CCATATATGTCACAACCCAGGAGGCAATGTCAGCTCTGTATGCAAGTGTGTCGTC 1155
Db 241 CCATATATGTCACAACCCAGGAGGCAATGTCAGCTCTGTATGCAAGTGTGTCGTC 300
QY 1156 CACACTGGTTTTAAATGGCCATGCTGCGCAATTACTTCTGTCTACATAAAGCTAGCAATGCG 1215
Db 301 CACACTGGTTTTAAATGGCCATGCTGCGCAATTACTTCTGTCTACATAAAGCTAGCAATGCG 360
QY 1216 CTCGTGATCAAAATGAATGACTCCATTTGATCTACAGTGATATTAGATCGGTACTCAGC 1275
Db 361 CTCGTGATCAAAATGAATGACTCCATTTGATCTACAGTGATATTAGATCGGTACTCAGC 420
QY 1276 CAACAAGCTATGTCCTTTTATATCAGTCCCATGATGTGAAAAATGGAGGTGAACTT 1335
Db 421 CAACAAGCTATGTCCTTTTATATCAGTCCCATGATGTGAAAAATGGAGGTGAACTT 480
QY 1336 ACTCATCCACCCATAGCCCGGCGCAGTCCCTCTCCCGCCCGCTCATCAGTCAGCGGGT 1395
Db 481 ACTCATCCACCCATAGCCCGGCGCAGTCCCTCTCCCGCCCGCTCATCAGTCAGCGGGT 540
QY 1396 GTCACCAACAAAGGCTGCGCCAGGCTTTATCGGACCAACAGCTTCCCTCTCACATGATA 1455
Db 541 GTCACCAACAAAGGCTGCGCCAGGCTTTATCGGACCAACAGCTTCCCTCTCACATGATA 600
QY 1456 AAGAATCCACTCACTTAATGGGACTGGACCATTTGAAGACAGCCAGCAGTTCCATG 1515
Db 601 AAGAATCCACTCACTTAATGGGACTGGACCATTTGAAGACAGCCAGCAGTTCCATG 660
QY 1516 TCGAGTCTTAACGGGAATTCAGTGTCAACAGGCTAGTCTGTTAATGCTTCA 1569
Db 661 TCGAGTCTTAACGGGAATTCAGTGTCAACAGGCTAGTCTGTTAATGCTTCA 714

RESULT 11
BM563326
LOCUS
DEFINITION AGENCOURT 6589448 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5441601
5', mRNA sequence.
ACCESSION BM563326
VERSION BM563326.1 GI:18810154
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Butheria; Primates; Catarrhini; Hominidae; Homo.

```

```

1 (bases 1 to 1088)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rgs@bbs.fda.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1916 row: c column: 10
High quality sequence stop: 611.
Location/Qualifiers
1..1088
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5441601"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="PH108 (phage-resistant)"
/clone_lib="NIH_MGC_98"
/note="Organ: Brain; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
ORIGIN
Query Match      25.0%; Score 604; DB 12; Length 1088;
Best Local Similarity 99.8%; Pred. No. 5.9e-298;
Matches 654; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 133 GCGAGCTCCGAGGCGAGTCTCACCTGGAGACATGGATGCGAGGTTCTGCCAGCTGGGGTCT 192
Db 1 GCGAGCTCCGAGGCGAGTCTCACCTGGAGACATGGATGCGAGGTTCTGCCAGCTGGGGTCT 60
QY 193 GTGTCTTCATTAATGATGTTGTTCAATCACACATCTTTTAGGACCACTGCTGCTGTCT 252
Db 61 GTGTCTTCATTAATGATGTTGTTCAATCACACATCTTTTAGGACCACTGCTGCTGTCT 120
QY 253 GTAGTTTATTCGAGTTCATCTGCTACCTGATTAATCAAAACCATCACCAAAAGATCAA 312
Db 121 GTAGTTTATTCGAGTTCATCTGCTACCTGATTAATCAAAACCATCACCAAAAGATCAA 180
QY 313 GCCCTAGGTGATGGCATCGCTCTCCACAGAAAGTCTTTTCCCATCTGAGAAGATTTGT 372
Db 181 GCCCTAGGTGATGGCATCGCTCTCCACAGAAAGTCTTTTCCCATCTGAGAAGATTTGT 240
QY 373 CTTAAGTGGCAAACTCATAGAGTTGGAGTGGGCTCCAGAAATTTGGCAATACCTGT 432
Db 241 CTTAAGTGGCAAACTCATAGAGTTGGAGTGGGCTCCAGAAATTTGGCAATACCTGT 300
QY 433 TTTGCCAATCAGCAGTGTGTTTAACTTACCTACACACCTCTTTGCCAATACATGCTA 492
Db 301 TTTGCCAATCAGCAGTGTGTTTAACTTACCTACACACCTCTTTGCCAATACATGCTA 360
QY 493 TCACATGAAACACCTCCAAAACATGTCATGCAAGAGGCTTTTGTATGATGTCAGATGCA 552
Db 361 TCACATGAAACACCTCCAAAACATGTCATGCAAGAGGCTTTTGTATGATGTCAGATGCA 420
QY 553 GCACATATTACCCAGGCACTCAGTAATCTCTGGGAGCGTTATTAAACCAATGTTGTCATC 612
Db 421 GCACATATTACCCAGGCACTCAGTAATCTCTGGGAGCGTTATTAAACCAATGTTGTCATC 480
QY 613 AATGAGATCGGCGTATAGCTAGGACCTCCGTTTGGAAACCAAGAGATGCCATGAA 672
Db 481 AATGAGATCGGCGTATAGCTAGGACCTCCGTTTGGAAACCAAGAGATGCCATGAA 540

```

QY	673	TTCCCTTCAATACACTGTTGATGCTATGAGAAAGCATCTTGAATGGAGCAATAAATTA	732
Db	541	TTCCCTTCAATACACTGTTGATGCTATGAGAAAGCATCTTGAATGGAGCAATAAATTA	600
QY	733	GACAGACACACCCAGCCACCACTCTTGTGTTGTCAGATATTTGGAGGATACCTAA	787
Db	601	GACAGACACACCCAGCCACCACTCTTGTGTTGTCAGATATTTGGAGGATACCTAA	655
RESULT 12			
LOCUS	BQ778581	595 bp mRNA linear EST 26-JUL-2002	
DEFINITION	il29f12.y1 HR85 islet Homo sapiens cDNA clone IMAGE:6031630 5', mRNA sequence.		
ACCESSION	BQ778581	GI:21987053	
VERSION			
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE			
AUTHORS	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 595)		
	Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wyllie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T., Jackson, Y. and Bowers, Y.		
	Endocrine Pancreas Consortium		
TITLE	Unpublished (2000)		
JOURNAL	Other ESTs: il29f12.x1		
COMMENT	Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue Endocrine Pancreas Consortium Harvard University, Howard Hughes Medical Institute Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138 Tel: 617-495-1812 Fax: 617-495-8557 Email: dmelton@hobp.harvard.edu Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu) Seq primer: -40RP from Gibco High quality sequence stop: 485.		
FEATURES			
source	Location/Qualifiers		
	1..595		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6031630"		
	/tissue_type="Purified pancreatic islet"		
	/lab_host="DH10B"		
	/clone_libs="HR85 islet"		
	/note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1: NotI; Site_2: XhoI; cDNA made by oligo-dr priming. Size-selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel: 314-362-1916, Fax: 314-747-2692."		
ORIGIN			
Query Match	24.3%; Score 585; DB 13; Length 595;		
Best Local Similarity	100.0%; Fred. No. 2.9e-288; Indels 0; Gaps 0;		
Matches 585; Conservative	0; Mismatches 0;		
QY	1371	CGCCCCCGTCATCAGTCAGCGGGTGTTCACCAACAAACAGCGTCGCCAGGCTTTATCGG	1430
Db	11	CGCCCCCGTCATCAGTCAGCGGGTGTTCACCAACAAACAGCGTCGCCAGGCTTTATCGG	70
QY	1431	ACCACAGCTTCCTCTCACATGATAAAGAAATCCACCTCACTTAAATGGACCTGGACCAATT	1490
Db	71	ACCACAGCTTCCTCTCACATGATAAAGAAATCCACCTCACTTAAATGGACCTGGACCAATT	130
QY	1491	GAAGACACGCCAGCAGTTCATGTCAGTCTTAACGGGAATTCAGTGTCAACAGGGC	1550
Db	131	GAAGACACGCCAGCAGTTCATGTCAGTCTTAACGGGAATTCAGTGTCAACAGGGC	190
QY	1551	TAGTCTGTGTAATGCTTCAGCTTCTGTCACAACTCGGTGAGTAAATAGTCTCCTCAGTGAT	1610
Db	191	TAGTCTGTGTAATGCTTCAGCTTCTGTCACAACTCGGTGAGTAAATAGTCTCCTCAGTGAT	250
QY	1611	CCGAGAACATCTTAAGAAACAAAAATTAACATCAGTATTCAACAAGTTCCTCTTCG	1670
Db	251	CCGAGAACATCTTAAGAAACAAAAATTAACATCAGTATTCAACAAGTTCCTCTTCG	310
QY	1671	CCAGTGTGAGTCTCAACCTAACCTTCATAGTAATTTCTTTGGAGAACCTTACCAAGCCCGT	1730
Db	311	CCAGTGTGAGTCTCAACCTAACCTTCATAGTAATTTCTTTGGAGAACCTTACCAAGCCCGT	370
QY	1731	TCCTCTTCTACCATTAACCAATTTCTGAGTACAGTCTACCTCGAACGCAATCTACGATGC	1790
Db	371	TCCTCTTCTACCATTAACCAATTTCTGAGTACAGTCTACCTCGAACGCAATCTACGATGC	430
QY	1791	AGTTTCTAGTAAAGTAACAAACCGATCCCGCAGTGAATCTCTGCTCCAGCCCGTGAT	1850
Db	431	AGTTTCTAGTAAAGTAACAAACCGATCCCGCAGTGAATCTCTGCTCCAGCCCGTGAT	490
QY	1851	GAATGGCAATCAAGCTGAATCTCCAGCTGTGCTGCTCCCTATGGCGCCGAGTCTCTGA	1910
Db	491	GAATGGCAATCAAGCTGAATCTCCAGCTGTGCTGCTCCCTATGGCGCCGAGTCTCTGA	550
QY	1911	GGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGGAGAAATGGAT	1955
Db	551	GGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGGAGAAATGGAT	595
RESULT 13			
LOCUS	AW575816	557 bp mRNA linear EST 15-MAR-2000	
DEFINITION	UI-HF-BNO-afg-g-10-0-UI.s1 NIH_MGC_50 Homo sapiens cDNA clone IMAGE:3066811 3', mRNA sequence.		
ACCESSION	AW575816		
VERSION	AW575816.1	GI:7247355	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 557)		
REFERENCE	NIH-MGC http://mgi.nci.nih.gov/.		
AUTHORS	National Institutes of Health, Mammalian Gene Collection (MGC)		
TITLE	Unpublished (1999)		
JOURNAL	Contact: Robert Strausberg, Ph.D.		
COMMENT	Email: cgapbs-remail.nih.gov Oligo-dr track not found, Not I site shown in beginning of sequence is likely internal to the message. Tissue Procurement: Louis M. Staudt, M.D., Ph.D. cDNA Library Preparation: M.B. Soares Lab cDNA Library Arrayed by: M.B. Soares Lab DNA Sequencing by: M.B. Soares Lab Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html Seq primer: M13 forward POLYA=No.		
FEATURES			
source	Location/Qualifiers		
	1..557		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3066811"		


```

RESULT 15
AW298129/c
LOCUS
DEFINITION
  UI-H-BW0-ajs-e-07-0-UI.sl NCI CGAP_Sub6 Homo sapiens cDNA clone
  IMAGE:2733012 3', mRNA sequence.
ACCESSION
  AW298129
VERSION
  EST.
SOURCE
  1 GI:6704765
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 538)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs@mail.nih.gov
  Oligo-dt track not found, Not 1 site shown in beginning of sequence
  is likely internal to the message. cDNA library preparation: M.B.
  Soares Lab Clone distribution: NCI-CGAP clone distribution
  information can be found through the I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/bbrp/image/image.html
  Seq primer: M13 Forward
  POLYA=NO.
  Location/Qualifiers
  1..538
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:2733012"
    /lab_host="DH10B (Life Technologies)"
    /note="Vector: pT73D-pac (Pharmacia) with a modified
    polylinker; Site 1: Not 1; Site 2: Eco RI; NCI CGAP Sub6
    is a subtracted library derived from BW, which consists of
    a mixture of four normalized libraries: NCI CGAP_Brn50,
    NCI CGAP_Lu13, NCI CGAP_Ov18, GBC1. The NCI CGAP Sub6
    library had 7 million recombinants. A single-stranded DNA
    preparation of BW was used as a tracer in a subtractive
    hybridization with a driver comprising: the IMAGE pool
    (NCI CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683,
    3798-3803 (IMAGE ClonesIDs 1323376-1323911,
    1456008-1456775, 1500552-1502855); NCI CGAP Kids pool 1
    LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE ClonesIDs
    1323912-1325831, 1471368-1472903, 1492104-1493255);
    NCI CGAP Lu5 pool 1 LLAM 3575-3582, 3851-3854 (IMAGE
    ClonesIDs 1414920-1417991, 1520904-1522439); NCI CGAP_GC4
    pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
    ClonesIDs 1257096-1258631, 1469064-1470983,
    1475592-1476743); NCI CGAP Pr22 pool 1 LLAM 2457-2459,
    2758-2759, 3062-3068 (IMAGE ClonesIDs 985608-986759,
    1101192-1101959, 1217928-1220615); NCI CGAP_Co10 pool 1
    LLAM 2644-2653, 2871-2872 (IMAGE ClonesIDs
    1057416-1061255, 1144584-1145351). (50% of the driver
    population), plus a pool of 3,840 arrayed clones from
    NCI CGAP Sub1 (IMAGE ClonesIDs 2708616-2710535) and
    NCI CGAP Sub2 (IMAGE ClonesIDs 2710536-2712455) (20% of
    the driver population), plus a pool of 11,136 clones from
    NCI CGAP Sub3 (IMAGE ClonesIDs 2712456-2723591) (30% of
    the driver population). Subtraction was performed as
    previously described (Honald, Lennon & Soares (1996):
    Normalization and Subtraction: Two Approaches To
    Facilitate Gene Discovery. Genome Research 6, 791-806.
    TAG TISSUE=brain
    TAG LIB=NCI CGAP_Brn50
    TAG_SEQ=TTTGG"

FEATURES
  source
    1..538
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:3077260"
      /tissue_type="lymph"

  Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 1834 TGCTCCAGCCGCTGATGAATGGCAAAATCCAGCTGAATCCAGCGTGTGGTGCCTAT 1893
  Db 538 TGCTCCAGCCGCTGATGAATGGCAAAATCCAGCTGAATCCAGCGTGTGGTGCCTAT 479
  QY 1894 GCGCGCGAGTCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGAGAAATGGG 1953
  Db 478 GCGCGCGAGTCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGAGAAATGGG 419
  QY 1954 ATTGGTACGATTGTGAGCTCCCACTCTCCCGGCCAAGATGCCAAGATGAGGAGGCACCT 2013
  Db 418 ATTGGTACGATTGTGAGCTCCCACTCTCCCGGCCAAGATGCCAAGATGAGGAGGCACCT 359
  QY 2014 CCGCAGGAGCTTCAGAAACCCATGACCTTAAACGGTGTCTAATAGTCACAGCAGCAGT 2073
  Db 358 CCGCAGGAGCTTCAGAAACCCATGACCTTAAACGGTGTCTAATAGTCACAGCAGCAGT 299
  QY 2074 GACCCGAAAGAAAACGGCTAGCGCTGTATGGTCCAGCTGCCAAGGCCAGCTGCCCTG 2133
  Db 298 GACCCGAAAGAAAACGGCTAGCGCTGTATGGTCCAGCTGCCAAGGCCAGCTGCCCTG 239
  QY 2134 CACTCAGAAAATCCCTTTGCTTAAGGCAACCGTCTTCCTGGAAGTTGATGCTGCTCT 2193
  Db 238 CACTCAGAAAATCCCTTTGCTTAAGGCAACCGTCTTCCTGGAAGTTGATGCTGCTCT 179
  QY 2194 TTGCTGTCTCTCCCAAGACACAAATCTTAGAGACCTTCAGGCTTAGCAACAACTGAA 2253
  Db 178 TTGCTGTCTCTCCCAAGACACAAATCTTAGAGACCTTCAGGCTTAGCAACAACTGAA 119
  QY 2254 GGCTCGACGATGAAATGAGTGCACCTGGAGCAGAGAGGGGCCCTCCCGAGGACCGGCAC 2313
  Db 118 GGCTCGACGATGAAATGAGTGCACCTGGAGCAGAGAGGGGCCCTCCCGAGGACCGGCAC 59
  QY 2314 GCGAGCCTCAGCTGCGAGCCCGCCCGCGAATCCTCGAGAGGACGAGATGCGGCG 2371
  Db 58 GCGAGCCTCAGCTGCGAGCCCGCCCGCGAATCCTCGAGAGGACGAGATGCGGCG 1

  RESULT 16
  AW500503
  LOCUS
  DEFINITION
    UI-HF-BN0-akj-a-03-0-UI.r1 NIH_MGC_50 Homo sapiens cDNA clone
    IMAGE:3077260 5', mRNA sequence.
  ACCESSION
    AW500503
  VERSION
    1 GI:7113179
  KEYWORDS
    EST.
  SOURCE
    Homo sapiens (human)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    1 (bases 1 to 542)
    NIH-MGC http://mgc.nci.nih.gov/.
    National Institutes of Health, Mammalian Gene Collection (MGC)
    Unpublished (1999)
    Contact: Robert Strausberg, Ph.D.
    Email: cgapbs@mail.nih.gov
    Eco RI site shown at the beginning of the sequence.
    Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
    cDNA Library Preparation: M.B. Soares Lab
    cDNA Library Arrayed by: M.B. Soares Lab
    DNA Sequencing by: M.B. Soares Lab
    Clone distribution: MGC clone distribution information can be
    found through the I.M.A.G.E. Consortium/LLNL at:
    www-bio.llnl.gov/bbrp/image/image.html
    Seq primer: M13 Forward.
    Location/Qualifiers
    1..542
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:3077260"
      /tissue_type="lymph"

```

ORIGIN

Query Match 22.3%; Score 538; DB 10; Length 538;
 . Best Local Similarity 100.0%; Pred. No. 3.8e-264;

/cell_type="germinal center B cells"
 /cell_line="WGC85"
 /lab_host="DH10B (LT1)"
 /clone_lib="NIH MGC 50"
 /note="Vector: pT73-Pac; Site 1: NotI; Site 2: Eco RI;
 Constructed from size fractionated cytoplasmic mRNA
 (3.5-4.4kb). Directionally cloned. Cells provided by
 Louis M. Staudt, Ph.D. Library preparation by Maria de
 Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

ORIGIN

Query Match 22.2%; Score 535; DB 10; Length 542;
 Best Local Similarity 100.0%; Pred. No. 1.3e-262;
 Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1265 CGGTACTACGCAACAGCCTATGCTCTTTATATACAGGTCCTCATGATGAAAATG 1324
 Db 8 CGGTACTACGCAACAGCCTATGCTCTTTATATACAGGTCCTCATGATGAAAATG 67

QY 1325 GAGGTGAACCTTACTCATCCACCATAGCCCGCCAGTCCTCTCCCGCCCGTCATCA 1384
 Db 68 GAGGTGAACCTTACTCATCCACCATAGCCCGCCAGTCCTCTCCCGCCCGTCATCA 127

QY 1395 GTCAGCGGTTGTACCAACAAACAGGTCGCCAGGCTTTATCGACACAGCTTCCT 1444
 Db 128 GTCAGCGGTTGTACCAACAAACAGGTCGCCAGGCTTTATCGACACAGCTTCCT 187

QY 1445 CTCACATGATAAGAAATCCACTCTTAATGAGACTGACCATGAAAGACAGCCAA 1504
 Db 188 CTCACATGATAAGAAATCCACTCTTAATGAGACTGACCATGAAAGACAGCCAA 247

QY 1505 GCAGTTCATGCGAGTCTCAACGGGAATTCAGAGTCAACAGGGCTAGTCCTGTAATG 1564
 Db 248 GCAGTTCATGCGAGTCTCAACGGGAATTCAGAGTCAACAGGGCTAGTCCTGTAATG 307

QY 1565 CTTGAGCTTCTGTCCAAATCGTGTAGTAAATGAGTCTCAGTGATCCAGAAATCCTTA 1624
 Db 308 CTTGAGCTTCTGTCCAAATCGTGTAGTAAATGAGTCTCAGTGATCCAGAAATCCTTA 367

QY 1625 AGAACAAATAATACATCAGTATTCACAAAGTTCGCTGTTGCCAGTGTAGTCTC 1684
 Db 368 AGAACAAATAATACATCAGTATTCACAAAGTTCGCTGTTGCCAGTGTAGTCTC 427

QY 1685 AACCTAACCTTCATAGTAAATCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCA 1744
 Db 428 AACCTAACCTTCATAGTAAATCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCA 487

QY 1745 TTACCAATTCGAGTACAGTCTACCTCGAAGCATCTACGATGTAGTTCCTAG 1799
 Db 488 TTACCAATTCGAGTACAGTCTACCTCGAAGCATCTACGATGTAGTTCCTAG 542

RESULT 17
 AW976266/c
 LOCUS AW976266 686 bp mRNA linear EST 02-JUN-2000
 DEFINITION EST388375 MAGE resequences, MAGN Homo sapiens cDNA, mRNA sequence.
 ACCESSION AW976266
 VERSION AW976266.1 GI:8167492
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 686)
 Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C.,
 Holt, I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and
 Quackenbush, J.

TITLE Assessment of gene expression patterns in a model of colon tumor
 metastasis using a 19,200 element cDNA microarray
 JOURNAL Unpublished (2000)
 COMMENT Contact: John Quackenbush
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 3528
 Fax: 301 838 0208
 Email: johnq@tigr.org
 Plate: 358
 Seq primer: Forward.
 Location/Qualifiers
 source
 1..686
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="MAGE resequences, MAGN"
 /note="Vector: pBluescriptSkM"

ORIGIN

Query Match 22.1%; Score 534; DB 10; Length 686;
 Best Local Similarity 99.8%; Pred. No. 4.5e-262;
 Matches 584; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1755 TGCAGTACAGTCTACCTCGAAGCGATCTACGATGTCAGTTCTAGTAAAGTAAACACC 1814
 Db 585 TGCAGTACAGTCTACCTCGAAGCGATCTACGATGTCAGTTCTAGTAAAGTAAACACC 526

QY 1815 GATCCCCCGCAGTGAATCTCTGCTCCAGCCCTGATGAATGCAATCCAAAGCTGAATC 1874
 Db 525 GATCCCCCGCAGTGAATCTCTGCTCCAGCCCTGATGAATGCAATCCAAAGCTGAATC 466

QY 1875 CAGCGTGTGTGTCCTATGCGCGCAGTCTCTCTGAGGACTCTGACGAGTCAAAAGG 1934
 Db 465 CAGCGTGTGTGTCCTATGCGCGCAGTCTCTCTGAGGACTCTGACGAGTCAAAAGG 406

QY 1935 GCTGGGCAAGGAGATGGAGTTGGTACGATGTGAGTCTCCACTCTCCGGCCAAAGATGC 1994
 Db 405 GCTGGGCAAGGAGATGGAGTTGGTACGATGTGAGTCTCCACTCTCCGGCCAAAGATGC 346

QY 1995 CGAAGTACAGGAGGCACTCCGACAGCTTCAAGAACCCATGACCCCTAAACCGTGTAA 2054
 Db 345 CGAAGTACAGGAGGCACTCCGACAGCTTCAAGAACCCATGACCCCTAAACCGTGTAA 286

QY 2055 TAGTCACAGACGACAGTACCCGAAAGAAACCGCCTAGCGCCTGATGTGCCAGCTG 2114
 Db 285 TAGTCACAGACGACAGTACCCGAAAGAAACCGCCTAGCGCCTGATGTGCCAGCTG 226

QY 2115 CCAAGGCCAGCTCCCTGCGACTCAGAAATCCCTTTGCTAAGGCAACCGTCTTCCTGG 2174
 Db 225 CCAAGGCCAGCTCCCTGCGACTCAGAAATCCCTTTGCTAAGGCAACCGTCTTCCTGG 166

QY 2175 AAAGTTGATGCTGCTCTCTTTGCTGCTCTCCAGAACACAAATCTTTAGAGACCTTCAG 2234
 Db 165 AAAGTTGATGCTGCTCTCTTTGCTGCTCTCCAGAACACAAATCTTTAGAGACCTTCAG 106

QY 2235 GCTTAGCAACAACTGAAGGCTCGACGGATGAATGATGCACCTGGAGCAGAGGGG 2294
 Db 105 GCTTAGCAACAACTGAAGGCTCGACGGATGAATGATGCACCTGGAGCAGAGGGG 46

QY 2295 CCCTCCCGAGGACCGCGACCGGAGCTTCAGCTTCAGCTGGAGCCCCCG 2339
 Db 45 CCCTCCCGAGGACCGCGACCGGAGCTTCAGCTTCAGCTGGAGCCCCCG 1

RESULT 18

AW936218/c
 LOCUS AW936218 533 bp mRNA linear EST 08-MAR-2000
 DEFINITION w63d05.x1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:246009 3',
 mRNA sequence.
 ACCESSION AW936218
 VERSION AW936218.1 GI:5675088
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 533)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert length: 863 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 490.

FEATURES
source
1. 533
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2460009"
/sex="male"
/tissue_type="normal prostate"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Pr22"
/notes="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from normal prostate bulk tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 22.1%; Score 533; DB 9; Length 533;
Best Local Similarity 100.0%; Pred. No. 1.4e-261;
Matches 533; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1840 CAGCCCGTGATGATGGCAATCAAGCTGAACCTCAGCGTGTCTGCGTGGTGGCCCTATGGCGCC 1899
DB 533 CAGCCCGTGATGATGGCAATCAAGCTGAACCTCAGCGTGTCTGCGTGGTGGCCCTATGGCGCC 474
QY 1900 GAGTCCTCTGAGACTCTGACAGGAGTCAAGGGGCTGGGCAAGAGATGGGATTGGT 1959
DB 473 GAGTCCTCTGAGACTCTGACAGGAGTCAAGGGGCTGGGCAAGAGATGGGATTGGT 414
QY 1960 ACGATTGTGAGCTCCCACTCTCCGGCCAAAGATGCCGAAGATGAGGAGGCCACTCCGCAC 2019
DB 413 ACGATTGTGAGCTCCCACTCTCCGGCCAAAGATGCCGAAGATGAGGAGGCCACTCCGCAC 354
QY 2020 GAGCTTCAAGAACCCATGACCTTAAACGGTCTAATAGTCAGACAGCGAGATGGATTGGT 2079
DB 353 GAGCTTCAAGAACCCATGACCTTAAACGGTCTAATAGTCAGACAGCGAGATGGATTGGT 294
QY 2080 AAGAAACCGCTAGCGCTGTGTCAGCTGCCAAGCGGCTGGCTGCTCTCTTTGCTG 2139
DB 293 AAGAAACCGCTAGCGCTGTGTCAGCTGCCAAGCGGCTGGCTGCTCTCTCTGCTCA 234
QY 2140 GAAATCCCTTTGTTAAGGCAACCGTCTTCCTGAAAGTTGATGCTGCTCTTTGCTG 2199
DB 233 GAAATCCCTTTGTTAAGGCAACCGTCTTCCTGAAAGTTGATGCTGCTCTTTGCTG 174
QY 2200 TCTCTCCAGAAGACAAATCTTAGACACTTACAGCTTACAGCTTACAGCAACAACTGAAAGGCTCG 2259
DB 173 TCTCTCCAGAAGACAAATCTTAGACACTTACAGCTTACAGCTTACAGCAACAACTGAAAGGCTCG 114
QY 2260 ACGATGAAATGAGTGCACCTGGACAGAGAGGGGCTCCCGAGGACCGGAGCCGAG 2319
DB 113 ACGATGAAATGAGTGCACCTGGACAGAGAGGGGCTCCCGAGGACCGGAGCCGAG 54
QY 2320 CCTCAGCCTGGCAGCCCGCGCCGGAATCCCTGGAGGAGCGAGATGGCGCGC 2372

Db 53 CCTAGCCTGGCAGCCCGCGCCGGAATCCCTGGAGAGCCAGATCGCGCCG 1
|||||
RESULT 19
AI738779/c 533 bp mRNA linear EST 21-DEC-1999
LOCUS wi36e04.x1 NCI_CGAP_Co16 Homo sapiens cDNA clone IMAGE:2392350 3',
DEFINITION mRNA sequence.
AI738779
AI738779.1 GI:5100760
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE 1 (bases 1 to 553)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert length: 661 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 507.
Location/Qualifiers
1. 553
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2392350"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Co16"
/notes="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI CGAP Co16 was prepared, and ss circles were made in vitro. Following HAP hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 22.1%; Score 533; DB 9; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.4e-261;
Matches 533; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1840 CAGCCCGTGATGATGGCAATCAAGCTGAACCTCAGCGTGTCTGCGTGGTGGCCCTATGGCGCC 1899
DB 533 CAGCCCGTGATGATGGCAATCAAGCTGAACCTCAGCGTGTCTGCGTGGTGGCCCTATGGCGCC 474
QY 1900 GAGTCCTCTGAGACTCTGACAGGAGTCAAGGGGCTGGGCAAGAGATGGGATTGGT 1959
DB 473 GAGTCCTCTGAGACTCTGACAGGAGTCAAGGGGCTGGGCAAGAGATGGGATTGGT 414
QY 1960 ACGATTGTGAGCTCCCACTCTCCGGCCAAAGATGCCGAAGATGAGGAGGCCACTCCGCAC 2019
DB 413 ACGATTGTGAGCTCCCACTCTCCGGCCAAAGATGCCGAAGATGAGGAGGCCACTCCGCAC 354
QY 2020 GAGCTTCAAGAACCCATGACCTTAAACGGTCTAATAGTCAGACAGCGAGATGGATTGGT 2079
DB 353 GAGCTTCAAGAACCCATGACCTTAAACGGTCTAATAGTCAGACAGCGAGATGGATTGGT 294

QY 2080 AAAGAAACGGCTAGCGCTGATGTGCGCAGCTGCCAAGGCGAGCGCTGCCCTGCACTCA 2139
 DB 293 AAAGAAACGGCTAGCGCTGATGTGCGCAGCTGCCAAGGCGAGCGCTGCCCTGCACTCA 234
 QY 2140 GAAATCCCTTTGCTTAAGGCAACGGCTTCTCTGGAAGCTTGATGCTCTCTTTGCTG 2199
 DB 233 GAAATCCCTTTGCTTAAGGCAACGGCTTCTCTGGAAGCTTGATGCTCTCTTTGCTG 174
 QY 2200 TCTCTCCCAAGACAAATCTTAGAGACCTTCAGGCTTAGCAAACTGAAAGGCTCG 2259
 DB 173 TCTCTCCCAAGACAAATCTTAGAGACCTTCAGGCTTAGCAAACTGAAAGGCTCG 114
 QY 2260 ACCGATGAATCAGTGCACCTGAGCAGAGAGGGCGCTCCCGAGGACCGCGACGCCGAG 2319
 DB 113 AGGATGAATCAGTGCACCTGAGCAGAGAGGGCGCTCCCGAGGACCGCGACGCCGAG 54
 QY 2320 CTTAGCTGGCAGCGCCCGCCGCGAATCCCTGGAGGAGCCAGATGCGGCCGC 2372
 DB 53 CTTAGCTGGCAGCGCCCGCCGCGAATCCCTGGAGGAGCCAGATGCGGCCGC 1
 RESULT 20
 CFI35585
 LOCUS
 DEFINITION UI-HF-BN0-amg-h-08-0-UI.r1 NIH MGC_50 Homo sapiens cDNA clone
 IMAGE:3090278 5', mRNA sequence.
 ACCESSION CFI35585
 VERSION CFI35585.1 GI:33251029
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRP, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Louis Staudt
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
 http://genome.uiowa.edu/distribution/humanfl.html
 Seq primer: pYX-5.
 Location/Qualifiers
 1..722
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3090278"
 /tissue_type="lymph"
 /cell_type="germinal center B cells"
 /cell_line="MGC85"
 /lab_host="DH10B (LTI)"
 /clone_lib="NIH MGC 50"
 /note="Vector: p7T3-Pac; Site 1: NotI; Site 2: Eco RI;
 Constructed from size fractionated cytoplasmic mRNA
 (3.5-4.4kb). Directionally cloned. Cells provided by
 Louis M. Staudt, Ph.D. Library preparation by Maria de
 Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

Query Match 21.5%; Score 518; DB 14; Length 722;
 Best Local Similarity 100.0%; Pred. No. 7.4e-254;
 Matches 518; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1118 GAGAGCCAAATCTTACCGTCTTTGATGTCAGTCTGCTGCTCCACACTGGTTTAAATGCGATG 1177
 DB 62 GAGAGCCAAATCTTACCGTCTTTGATGTCAGTCTGCTGCTCCACACTGGTTTAAATGCGATG 121
 QY 1178 CTGGGCATTTACTTCTGTACATATAAAAGCTAGCAATGGCCCTCTGGTATCAATGAATGACT 1237
 DB 122 CTGGGCATTTACTTCTGTACATATAAAAGCTAGCAATGGCCCTCTGGTATCAATGAATGACT 181
 QY 1238 CCAATTGTATCTACCAAGTATATAGATCGGTACTCAGCCAAAGCCTATGTGCTCTTTT 1297
 DB 182 CCAATTGTATCTACCAAGTATATAGATCGGTACTCAGCCAAAGCCTATGTGCTCTTTT 241
 QY 1298 ATATCAGGTCCCATCATGTGAAATATGGAGGTGAATCTACTCATCCACCCATAGCCCG 1357
 DB 242 ATATCAGGTCCCATCATGTGAAATATGGAGGTGAATCTACTCATCCACCCATAGCCCG 301
 QY 1358 GCCAGTCTCTCCCGCCCGCGTCATCAGTCAGCGGGTGTTCACCAACAAAGGCTGCGC 1417
 DB 302 GCCAGTCTCTCCCGCCCGCGTCATCAGTCAGCGGGTGTTCACCAACAAAGGCTGCGC 361
 QY 1418 CAGGCTTTATCGGACCACAGCTTCCCTCTCATCATGATAAAGAAATCCACTCTTAAATG 1477
 DB 362 CAGGCTTTATCGGACCACAGCTTCCCTCTCATCATGATAAAGAAATCCACTCTTAAATG 421
 QY 1478 GGACTGGACCAATTGAAGACACAGCCCAAGCAGTTCATCGAGTCCGTAACGGGAATTCGA 1537
 DB 422 GGACTGGACCAATTGAAGACACAGCCCAAGCAGTTCATCGAGTCCGTAACGGGAATTCGA 481
 QY 1538 GTGTCAACAGGGCTAGTCTCTTAAATGCTTCTGCTTCTGTCACAACTGGTCAGTTAATA 1597
 DB 482 GTGTCAACAGGGCTAGTCTCTTAAATGCTTCTGCTTCTGTCACAACTGGTCAGTTAATA 541
 QY 1598 GGTCTCTAGTATCCAGAACATCTTAAGAAACAAAA 1635
 DB 542 GGTCTCTAGTATCCAGAACATCTTAAGAAACAAAA 579

RESULT 21

CFI35585
 LOCUS
 DEFINITION ir68901.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6607849 3',
 mRNA sequence.

ACCESSION CFI35585

VERSION CFI35585.1 GI:26675619

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2000)
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@bioh.p.harvard.edu
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Dr. Hiroshi Inoue

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 771)
NIH-MGC http://mgc.ncl.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11502 row: d column: 02
High quality sequence start: 2
High quality sequence stop: 766.
Location/Qualifiers
1..771
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5200465"
/lab_host="DH10B"
/clone_lib="NIH MGC 114"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: EcoRV (destroyed); RNA source anonymous pool of 6
male brains, age range 23-27 yo. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.5 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 019. Note:
this is a NIH_MGC Library."

FEATURES
source
Query Match 21.3%; Score 514; DB 14; Length 514;
Best Local Similarity 100.0%; Pred. No. 7.8e-252; Mismatches 0; Indels 0; Gaps 0;
Matches 514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1823 GCAGTGAATCTCTCCAGCCGCTGATGAATGGCAATCCAAAGCTCAACTCAGCGTGC 1882
Db 514 GCAGTGAATCTCTCCAGCCGCTGATGAATGGCAATCCAAAGCTCAACTCAGCGTGC 455
QY 1883 TGGTGCCTATGCGCGAGTCTCTGAGACTCTGACGAGGAGTCAAGGGGCTGGCA 1942
Db 454 TGGTGCCTATGCGCGAGTCTCTGAGACTCTGACGAGGAGTCAAGGGGCTGGCA 395
QY 1943 AGGAGATGGATGGTACGATGTGAGTCCCACTCTCCCGCCCAAGATGCCCAAGATG 2002
Db 394 AGGAGATGGATGGTACGATGTGAGTCCCACTCTCCCGCCCAAGATGCCCAAGATG 335
QY 2003 AGGAGCCACTCCGACAGCTTTCAAGAACCCATGACCTTAAACGGTGTAAATAGTCAG 2062
Db 334 AGGAGCCACTCCGACAGCTTTCAAGAACCCATGACCTTAAACGGTGTAAATAGTCAG 275
QY 2063 ACAGGACAGTACCCGAAGAAAACGGCTAGGCTGATGTGCGAGTGCAGTGCAGGCC 2122
Db 274 ACAGGACAGTACCCGAAGAAAACGGCTAGGCTGATGTGCGAGTGCAGTGCAGGCC 215
QY 2123 AGCTGCGCTGCACTCAGAAAATCCCTTTGCTAAGGCAACCGGTCTTCTCGGAAAGTTGA 2182
Db 214 AGCTGCGCTGCACTCAGAAAATCCCTTTGCTAAGGCAACCGGTCTTCTCGGAAAGTTGA 155
QY 2183 TGGCTGCTCTTGTCTGTCTCTCCAGAGACAAAATCTTAGAGACCTTCAGGGCTTAGCA 2242
Db 154 TGGCTGCTCTTGTCTGTCTCTCCAGAGACAAAATCTTAGAGACCTTCAGGGCTTAGCA 95
QY 2243 ACAACTGAAGGCTCAGCGATGAATGATGACCTGGAGCAGAGGGGCGCTCCCG 2302
Db 94 ACAACTGAAGGCTCAGCGATGAATGATGACCTGGAGCAGAGGGGCGCTCCCG 35
QY 2303 AGGACCGGACGCGGAGCTCAGCGCTGGCAGCCC 2336
Db 34 AGGACCGGACGCGGAGCTCAGCGCTGGCAGCCC 1

ORIGIN
Query Match 21.3%; Score 514; DB 12; Length 771;
Best Local Similarity 100.0%; Pred. No. 8.5e-252; Mismatches 0; Indels 0; Gaps 0;
Matches 514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 832 TTTGATCCATATCTTGATATAACATTGGAGATAAAGCTGCTCAGAGTGTCAACAGGCA 891
Db 142 TTTGATCCATATCTTGATATAACATTGGAGATAAAGCTGCTCAGAGTGTCAACAGGCA 201
QY 892 TTGAGCAGTTTGTGAAGCCGGAACAGCTTGATGGAGAAAACCTGTAACAAGTGCAGCAAG 951
Db 202 TTGAGCAGTTTGTGAAGCCGGAACAGCTTGATGGAGAAAACCTGTAACAAGTGCAGCAAG 261
QY 952 TGTAAAAGATGGTTCCAGCTTCAAGAGGTTCACTATCCATAGATCCTCTAATGTCTT 1011
Db 262 TGTAAAAGATGGTTCCAGCTTCAAGAGGTTCACTATCCATAGATCCTCTAATGTCTT 321
QY 1012 ACATTTCTCTGAAACGTTTTTGGAAAATTTACCGGTGAAAAAATGCTAAGGATGTGAA 1071
Db 322 ACATTTCTCTGAAACGTTTTTGGAAAATTTACCGGTGAAAAAATGCTAAGGATGTGAA 381
QY 1072 TACCTCAGTATCTTGATATTCGGCCATATATGTCTCAACCCACGAGAGCCAAATGTG 1131
Db 382 TACCTCAGTATCTTGATATTCGGCCATATATGTCTCAACCCACGAGAGCCAAATGTG 441
QY 1132 TACCTCAGTATCTTGATATTCGGCCATATATGTCTCAACCCACGAGAGCCAAATGTG 1191
Db 442 TACCTCAGTATCTTGATATTCGGCCATATATGTCTCAACCCACGAGAGCCAAATGTG 501
QY 1192 TGCTACATAAAGCTAGCAATGGCTCTGGTATCAATGAATGATGATCCTATCTATCTACC 1251
Db 502 TGCTACATAAAGCTAGCAATGGCTCTGGTATCAATGAATGATGATCCTATCTATCTACC 561
QY 1252 AGTGATATTAGATCGGTACTCAGCCCAACAGCCCTATGTGCTCTTTTATATCAGGTCCCAT 1311
Db 562 AGTGATATTAGATCGGTACTCAGCCCAACAGCCCTATGTGCTCTTTTATATCAGGTCCCAT 621

RESULT 22
BI755971 771 bp mRNA linear EST 25-SEP-2001
LOCUS 603030302P1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5200465 5',
DEFINITION mRNA sequence.
ACCESSION BI755971
VERSION BI755971.1 GI:15747549
KEYWORDS EST.
SOURCE Homo sapiens (human)

```

QY 1312 GATGTGAAAAATGAGGTGAACCTTACTCATCCCA 1345
|||||
Db 622 GATGTGAAAAATGAGGTGAACCTTACTCATCCCA 655

RESULT 23
AI472304/c
LOCUS
DEFINITION
tj87a05.x1 Soares NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:2148464 3', mRNA sequence.
AI472304
AI472304.1 GI:4334394
VERSION
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 488)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1221 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 464.
Location/Qualifiers
1..488
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2148464"
/lab_host="DH10B"
/clone_lib="Soares NSF_F8_9W_OT_PA_P_S1"
/notes="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NBHSF pool 1:
309384-310919, 323208-325895 Soares NB2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares NB2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NBHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 20.2%; Score 488; DB 9; Length 488;
Best Local Similarity 100.0%; Pred. No. 1.7e-238;
Matches 488; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1876 AGCGTGTGTGGCCCTATGGCGCGAGTCTCTCTGAGGACTCTGACGAGGAGTCAAGGGG 1935
|||||
Db 488 AGCGTGTGTGGCCCTATGGCGCGAGTCTCTCTGAGGACTCTGACGAGGAGTCAAGGGG 429

QY 1936 CTGGGCAAGAGATGGATGGTACGATTTGTGAGTCCCACTCTCCGGCCAGATGCC 1995
|||||
Db 428 CTGGGCAAGAGATGGATGGTACGATTTGTGAGTCCCACTCTCCGGCCAGATGCC 369

QY 1996 GAAGATGAGGAGGCCACTCCGCACGAGCTTCAGAACCCATGACCTTAAAGGTGCTAAT 2055
|||||
Db 368 GAAGATGAGGAGGCCACTCCGCACGAGCTTCAGAACCCATGACCTTAAAGGTGCTAAT 309

QY 2056 AGTGCACAGCAGCAGTGTAGTACCCGAAAGAAACGGCCTAGCGCTGATGGTCCAGTGC 2115

```

```

Db 308 AGTGCACAGCAGCAGTGTAGTACCCGAAAGAAACGGCCTAGCGCTGATGGTCCAGTGC 249
|||||
QY 2116 CAAGGCCAGCCTGCTGCACTCAGAAAATCCCTTTGCTAAGGCAAAACGGTCTTCTCTGGA 2175
|||||
Db 248 CAAGGCCAGCCTGCTGCACTCAGAAAATCCCTTTGCTAAGGCAAAACGGTCTTCTCTGGA 189
|||||
QY 2176 AAGTTGATGCTGCTCTCTCTTGTGTCTCTCCAGAAAGACAAAATCTTAGAGACTTCCAGG 2235
|||||
Db 188 AAGTTGATGCTGCTCTCTTGTGTCTCTCCAGAGAGACAAAATCTTAGAGACTTCCAGG 129
|||||
QY 2236 CTTAGCAACAACCTGAAAGCTCGACGATGAATGAGTGACCTCGAGCAGAGAGGGC 2295
|||||
Db 128 CTTAGCAACAACCTGAAAGCTCGACGATGAATGAGTGACCTCGAGCAGAGAGGGC 69
|||||
QY 2296 CTCTCCGAGGACCGCGACCGAGCTCGAGCTGAGCGCCCGCCCGCCGGAATCCCTGGAG 2355
|||||
Db 68 CTCTCCGAGGACCGCGACCGAGCTCGAGCTGAGCGCCCGCCCGCCGGAATCCCTGGAG 9
|||||
QY 2356 GAGCCAGA 2363
|||||
Db 8 GAGCCAGA 1

RESULT 24
AI309755/c
LOCUS
DEFINITION
qo69h09.x1 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1913825 3',
mRNA sequence.
AI309755
AI309755.1 GI:4004626
VERSION
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 465)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 870 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 433.
Location/Qualifiers
1..465
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1913825"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Kid5"
/notes="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
AAGTGAAGAAATCGCGCGCAATATTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

```

```

Query Match      19.3%; Score 465; DB 9; Length 465;
Best Local Similarity 100.0%; Pred. No. 1.1e-226;
Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1908 TGAGGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGGAGATGGATTGGTACGATTGT 1967
      |||
465  TGAGGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGGAGATGGATTGGTACGATTGT 406
      |||

1968 GAGCTCCCACTCTCCCGGCAAGATGCCAAGATGAGGAGGCACTCCCGACGAGTTCA 2027
      |||
405  GAGCTCCCACTCTCCCGGCAAGATGCCAAGATGAGGAGGCACTCCCGACGAGTTCA 346
      |||

2028 AGAACCACATGACCTAAACCGTGTCTAATAGTCAGACACGACGAGTACCGGAAGAAA 2087
      |||
345  AGAACCACATGACCTAAACCGTGTCTAATAGTCAGACACGACGAGTACCGGAAGAAA 286
      |||

2088 CGGCTAGGCTGTGATGCTGCGAGCTGCCAAGCCAGCCCTGCCCTGCACCTCAGAAATCC 2147
      |||
285  CGGCTAGGCTGTGATGCTGCGAGCTGCCAAGCCAGCCCTGCCCTGCACCTCAGAAATCC 226
      |||

2148 CTTTGTCTAGGCAAAACGGTCTTCTCGAAAAGTTGATGCCCTGCTCTCTCTCCC 2207
      |||
225  CTTTGTCTAGGCAAAACGGTCTTCTCGAAAAGTTGATGCCCTGCTCTCTCTCCC 166
      |||

2208 AGAAGCAAAATCTTAGAGACCTTCAGGCTTAGCAACAACTGAAGGCTCGACGATGA 2267
      |||
165  AGAAGCAAAATCTTAGAGACCTTCAGGCTTAGCAACAACTGAAGGCTCGACGATGA 106
      |||

2268 ANTAGTGCACCTGGAGCAGAGAGGGGCTCCCGAGGACCGGAGCCGAGCTCAGCC 2327
      |||
105  ANTAGTGCACCTGGAGCAGAGAGGGGCTCCCGAGGACCGGAGCCGAGCTCAGCC 46
      |||

2328 TGCAGCGCCCGCGCGAAATCCCTTGGAGGAGCCAGATCGCGCGCG 2372
      |||
45  TGCAGCGCCCGCGCGAAATCCCTTGGAGGAGCCAGATCGCGCGCG 1

RESULT 25
B0778360/c      516 bp      mRNA      linear      EST 26-JUL-2002
LOCUS          1129f12.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6031630 3',
DEFINITION      mRNA sequence.
ACCESSION      B0778360
VERSION        B0778360.1 GI:21986832
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
                1 (bases 1 to 516)
                Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
                Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
                Hillier, L., Narra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,
                Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J.,
                Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarelshvili, R.,
                Williams, T., Jackson, Y., and Bowers, Y.
                Endocrine Pancreas Consortium
                Unpublished (2000)
                Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
                Endocrine Pancreas Consortium
                Harvard University, Howard Hughes Medical Institute
                Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
                MA 02138
                Tel: 617-495-1812
                Fax: 617-495-8557
                Email: dmelton@biochem.harvard.edu
                Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
                Washington University Genome Sequencing Center For information on
                obtaining a clone please contact: Dr. Hiroshi Inoue
                (hinoue@im.wustl.edu)
                Seq primer: -40UP from Gibco
                High quality sequence stop: 439.

```

```

FEATURES
source
Location/Qualifiers
1..516
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6031630"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_1_b="HR85 islet"
/Note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
NotI; Site_2: XhoI; cDNA made by oligo-dT priming. 5'
Size selected on agarose gel. Average insert size -1kb. 5'
XhoI site was destroyed after directional cloning.
Amplified once. Contact information: Hiroshi Inoue, MD,
Metabolism Div. (Alan Permutt Lab), Washington University
School of Medicine, Box 8127, 660 South Euclid Ave., St.
Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
314-362-1916, Fax: 314-747-2692."

ORIGIN
Query Match      19.3%; Score 465; DB 13; Length 516;
Best Local Similarity 99.8%; Pred. No. 1.1e-226;
Matches 515; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1824 CAGTGAATCTGCTCCAGCCCGTGTATGATGCAATCCAACTCCAGCGTGT 1883
      |||
516  CAGTGAATCTGCTCCAGCCCGTGTATGATGCAATCCAACTCCAGCGTGT 457
      |||

1884 GGTGCCCTATGGCGCGAGTCCCTCTGAGGACTCTGAGGAGTCAAGGGGCTGGGCAA 1943
      |||
456  GGTGCCCTATGGCGCGAGTCCCTCTGAGGACTCTGAGGAGTCAAGGGGCTGGGCAA 397
      |||

1944 GGAGATGGGATTGGTACGATTGTGAGCTCCCACTCTCCCGGCCAAGATGCCGAAGATGA 2003
      |||
396  GGAGATGGGATTGGTACGATTGTGAGCTCCCACTCTCCCGGCCAAGATGCCGAAGATGA 337
      |||

2004 GGAGGCGCACTCCGCAAGAGCTTCAAGAACCCATGACCTTAAACGGTCTTAATAGTCAGA 2063
      |||
336  GGAGGCGCACTCCGCAAGAGCTTCAAGAACCCATGACCTTAAACGGTCTTAATAGTCAGA 277
      |||

2064 CAGGACAGTACCCGCAAGAAACGGCTAGCGCTGATGGTCCAGCTGCCAAGGCCA 2123
      |||
276  CAGGACAGTACCCGCAAGAAACGGCTAGCGCTGATGGTCCAGCTGCCAAGGCCA 217
      |||

2124 GCCTGCCCTGCACCTCAGAAAATCCCTTTGCTAAGGCAACGGTCTTCTTGAAAAGTTGAT 2183
      |||
216  GCCTGCCCTGCACCTCAGAAAATCCCTTTGCTAAGGCAACGGTCTTCTTGAAAAGTTGAT 157
      |||

2184 GCCTGCTCCTTTGCTGTCTCTCCAGAAACAAAATCTTAGAGACCTTCAGGCTTAGCAA 2243
      |||
156  GCCTGCTCCTTTGCTGTCTCTCCAGAAACAAAATCTTAGAGACCTTCAGGCTTAGCAA 97
      |||

2244 CAAACTGAAGGCTTGACGGATGAATCAGTGCACTGGAGCAGAGAGGGCCCTCCCGA 2303
      |||
96  CAAACTGAAGGCTTGACGGATGAATCAGTGCACTGGAGCAGAGAGGGCCCTCCCGA 37
      |||

2304 GGACCGGACCGCCGAGCTCAGCCCTGCGAGCCCGCCG 2339
      |||
36  GGACCGGACCGCCGAGCTCAGCCCTGCGAGCCCGCCG 1
      |||

RESULT 26
BUI48710
LOCUS          BUI48710
DEFINITION     AGENCOURT 8684473 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:6377023
                5', mRNA sequence.
ACCESSION      BUI48710
VERSION        BUI48710.1 GI:22662242
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE	1 (bases 1 to 988)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ ;
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LfCM2560 row: c column: 08 High quality sequence stop: 471.
FEATURES	Location/Qualifiers
source	1..988
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="IMAGE:6377023"
	/issue_type="neuroblastoma, cell line"
	/lab_host="DH10B (phage-resistant)"
	/clone_lib="NIH_MGC_4"
	/note="Organ: Brain; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
ORIGIN	
Query Match	19.1%; Score 460; DB 13; Length 988;
Best Local Similarity	99.6%; Pred. No. 4.7e-224;
Matches 560; Conservative	0; Mismatches 2; Indels 0; Gaps 0
QY	515 GTCATGCAGAAAGCGCTTTTGTATGATGTGTACAATGCAGACACATATTACCCAGGCACTCA 574
Db	33 GTCATGCAGAAAGCGCTTTTGTATGATGTGTACAATGCAGACACATATTACCCAGGCACTCA 92
QY	575 GTAATCCTGGGAGCTTATTAAACCAATTTGTTCATCAATGAGATGCGGGTAGTACCTA 634
Db	93 GTAATCCTGGGAGCTTATTAAACCAATTTGTTCATCAATGAGATGCGGGTAGTACCTA 152
QY	635 GGCACCTCCGTTTGTGGAAACCAAGAGATGCCATGAATTCCTTCATACACTGTTGATG 694
Db	153 GGCACCTCCGTTTGTGGAAACCAAGAGATGCCATGAATTCCTTCATACACTGTTGATG 212
QY	695 CTATGCAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCACCA 754
Db	213 CTATGCAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCACCA 272
QY	755 CTCTGTTTGTCCAGATATTTGGAGGATACCTAAGATCTAGAGTCAAAATGTTTAAATTGCA 814
Db	273 CTCTGTTTGTCCAGATATTTGGAGGATACCTAAGATCTAGAGTCAAAATGTTTAAATTGCA 332
QY	815 AGGGCGTTTCAGATACCTTTTGATCCATATCTTGATATATAAATTGGAGATAAAGGCTGCTC 874
Db	333 AGGGCGTTTCAGATACCTTTTGATCCATATCTTGATATATAAATTGGAGATAAAGGCTGCTC 392
QY	875 AGAGTGTCAACAGGCAATTTGGAGCAGTTTGTGAAGCCGGAACAGCTTGATGGAGAAACT 934
Db	393 AGAGTGTCAACAGGCAATTTGGAGCAGTTTGTGAAGCCGGAACAGCTTGATGGAGAAACT 452
QY	935 CGTACAGTGCAGCAAGTGTAAATAAGATGGTTCAGCTTCAAAAGAGGTTTCACATATCCATA 994
Db	453 CGTACAGTGCAGCAAGTGTAAATAAGATGGTTCAGCTTCAAAAGAGGTTTCACATATCCATA 512
QY	995 GATCCTCTAATGTTCTTACACTTCTCTGAAACGTTTTTGCAAAATTTTACCGGTGGAAAAA 1054

Db 513 GATCTCTAATGTTCTTACACTTTCTCTGAAACGTTTTCGCAATTTTACCGGTGAAAAA 572

QY 1055 TTGCTAAGGATGTGAAATACCC 1076
 |||||
 |||||

Db 573 TTGCTAAGGATGTGAAATACCC 594
 |||||
 |||||

RESULT 27
 AI081335/c

LOCUS
 DEFINITION ox22h05.x1 Soares_NhMPu_S1 Homo sapiens cDNA clone IMAGE:1661913
 3', mRNA sequence.

ACCESSION
 VERSION AI081335
 KEYWORDS AI081335.1 GI:3418127
 EST.

SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1 (bases 1 to 446)
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)

JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 869 Std Error: 0.00
 Seq primer: -40ml3 fwd. RT from Amersham
 High quality sequence stop: 422.

FEATURES
 source
 1..446
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1661913"
 /tissue_type="Pooled human melanocyte, fetal heart, and
 pregnant uterus"
 /lab_host="DH10B"
 /clone_lib="Soares_NhMPu_S1"
 /notes="Organ: mixed (see below); Vector: pT7T3D-Pac
 (Pharmacia) with a modified polylinker; Site_1: Not I;
 Site_2: Eco RI; Equal amounts of plasmid DNA from three
 normalized libraries (melanocyte 2N5HM, pregnant uterus
 N5HPU, and fetal heart N5HH19W) were mixed, and ss circles
 were made in vitro. Following HAP purification, this DNA
 was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from pools of
 5,000 clones made from the same 3 libraries. The pools
 consisted of I.M.A.G.E. clones 260232-265223,
 340488-345479, and 484488-489479."

	Query Match	18.5%;	Score 446;	DB 9;	Length 446;
	Best Local Similarity	100.0%;	Pred. No. 6e-217;		
	Matches 446;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1927	TCAAGGGGCTGGCGAAGCAGAGATGGATGGTACGATTGTGAGCTCCCACTCTCCGGC			1986
Db	446	TCAAGGGGCTGGCGAAGCAGAGATGGATGGTACGATTGTGAGCTCCCACTCTCCGGC			387
QY	1987	CAAGATGCCGAAGATGAGGAGGCCACTCCGCACGAGCTTCAAGAACCCATGACCTAAAC			2046
Db	386	CAAGATGCCGAAGATGAGGAGGCCACTCCGCACGAGCTTCAAGAACCCATGACCTAAAC			327
QY	2047	GGTGCTAATAGTGCACAGCGACAGTGCACCGCAAGAAAACGGCTACGCCGTGATGGT			2106
Db	326	GGTGCTAATAGTGCACAGCGACAGTGCACCGCAAGAAAACGGCTACGCCGTGATGGT			267
QY	2107	GCCAGCTGCCAAGGCGAGCCTGCCCTGCACCTCAGAAAATCCCTTGTCTAAGGCAACGGT			2166
Db	266	GCCAGCTGCCAAGGCGAGCCTGCCCTGCACCTCAGAAAATCCCTTGTCTAAGGCAACGGT			207

2167 CTCTCTGAAAGTTGATCGCTGCTCTTTTGTCTCTCTCCAGAGACAAATCTTAGAG 2226
 206 CTCTCTGAAAGTTGATCGCTGCTCTTTTGTCTCTCTCCAGAGACAAATCTTAGAG 147
 2227 ACCTTCAGGCTTAGCAACAACTGAAGGCTCGACGGATGAATGAGTGCACCTGGAGCA 2286
 146 ACCTTCAGGCTTAGCAACAACTGAAGGCTCGACGGATGAATGAGTGCACCTGGAGCA 87
 2287 GAGAGGGCCCTCCGAGGACCGGACGCGCTCTAGCCCTGGCAGCCCGCCCGCGAA 2346
 86 GAGAGGGCCCTCCGAGGACCGGACGCGCTCTAGCCCTGGCAGCCCGCCCGCGAA 27
 2347 TCCTCTGGAGGACCGAGATGCGGCCGC 2372
 26 TCCTCTGGAGGACCGAGATGCGGCCGC 1

BI914321 625 bp mRNA linear EST 17-OCT-2001
 603180996F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5245295 5',
 mRNA sequence.
 BI914321 GI:16198830
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-f@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM11618 row: o column: 24
 High quality sequence stop: 620.

FEATURES
 source
 1..625
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clones="IMAGE:5245295"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_121"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: EcoRV (destroyed); RNA source anonymous pool of 3
 fetal brains, female age 20 weeks, female age 24 weeks,
 and male age 26 weeks. Library is oligo-dT primed and
 directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.7 kb, insert size range
 0.7-3.5 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 017. Note:
 this is a NIH_MGC Library."

ORIGIN
 Query Match 18.4%; Score 444; DB 12; Length 625;
 Best Local Similarity 100.0%; Pred. No. 6.9e-216;
 Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

88 AAAGCTTCTGAATCTTCAGACCCATCAGCCTATCAGATTCAGCTGGCAGCTCCGAGCA 147
 63 AAAGCTTCTGAATCTTCAGACCCATCAGCCTATCAGATTCAGCTGGCAGCTCCGAGCA 122

148 GTCTCACCTGGAGACATGGATGCGAGTTCTGCCAGCTGGGTCGTGTCTTTCATTGAAT 207
 123 GTCTCACCTGGAGACATGGATGCGAGTTCTGCCAGCTGGGTCGTGTCTTTCATTGAAT 182
 208 GATGTCCTCAATCACAACATCTTTTAGGACGAGTACCTGGTGTCTAGTTTATTCGAGT 267
 183 GATGTCCTCAATCACAACATCTTTTAGGACGAGTACCTGGTGTCTAGTTTATTCGAGT 242
 268 TCATCTGTACCTGATAAATCAAAACCATCACCACAAAAGGATCAAGCCCTAGGTGATGC 327
 243 TCATCTGTACCTGATAAATCAAAACCATCACCACAAAAGGATCAAGCCCTAGGTGATGC 302
 328 ATCGCTCTCTCCACAGAAAGTTCTTTTCCCATCTGAGAAGATTGTCTTAAGTGGCAACA 387
 303 ATCGCTCTCTCCACAGAAAGTTCTTTTCCCATCTGAGAAGATTGTCTTAAGTGGCAACA 362
 388 ACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTTCCCATGACGA 447
 363 ACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTTCCCATGACGA 422
 448 CTGCGAGTGTAAACCTTACACACACCTCTTGGCAATTACATGCTATCAGATGAACACTCC 507
 423 CTGCGAGTGTAAACCTTACACACACCTCTTGGCAATTACATGCTATCAGATGAACACTCC 482
 508 AAAACATGTCATGCGAAGGCTTT 531
 483 AAAACATGTCATGCGAAGGCTTT 506

RESULT 29
 BQ956351
 LOCUS
 DEFINITION
 5', mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-f@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1CM2617 row: g column: 19
 High quality sequence stop: 473.

FEATURES
 source
 1..1085
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clones="IMAGE:6429354"
 /tissue_type="epidermoid carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_101"
 /note="Organ: lung; Vector: pOTB7; Site 1: EcoRI; Site 2:
 XhoI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."

ORIGIN

```

Query Match      18.1%; Score 436; DB 13; Length 1085;
Best Local Similarity 99.6%; Pred. No. 1e-211;
Matches 536; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 431 GTTTTCCCAATGAGCAGTGCAGTGTAACTTACCTACACACACCTCTTTGCCAATTTACATGC 490
DB 11 GTTTTCCCAATGAGCAGTGCAGTGTAACTTACCTACACACACCTCTTTGCCAATTTACATGC 70
QY 491 TATCATCATGAACACTCCAAACATGTCATGCAGAAAGCTTTTGTATGATGTGACAAATGC 550
DB 71 TATCATCATGAACACTCCAAACATGTCATGCAGAAAGCTTTTGTATGATGTGACAAATGC 130
QY 551 AAGCACAATATACCCAGGCACTCAGTAATCTTGGGAGCTTTATTAACCAATGTTTGTCA 610
DB 131 AAGCACAATATACCCAGGCACTCAGTAATCTTGGGAGCTTTATTAACCAATGTTTGTCA 190
QY 611 TCAATGAGATGGGGTATAGTAGGCACTCCGTTTGGAAACCAAGAGATGCCCATG 670
DB 191 TCAATGAGATGGGGTATAGTAGGCACTCCGTTTGGAAACCAAGAGATGCCCATG 250
QY 671 AATTCCTTCAATACACTGTTGATGTCATGCAGAAAGCATGCTTGAATGCGACCAATAAT 730
DB 251 AATTCCTTCAATACACTGTTGATGTCATGCAGAAAGCATGCTTGAATGCGACCAATAAT 310
QY 731 TAGACAGACACACCCAGGCGCACCTCTTGTGTCAGATATTTGGAGATACCTTAAGT 790
DB 311 TAGACAGACACACCCAGGCGCACCTCTTGTGTCAGATATTTGGAGATACCTTAAGT 370
QY 791 CTAGAGTCAATGTTTAAATGCAAGGGCGTTTCAGATACCTTTGATCCATATCTTGATA 850
DB 371 CTAGAGTCAATGTTTAAATGCAAGGGCGTTTCAGATACCTTTGATCCATATCTTGATA 430
QY 851 TAACTTGGAGATAAAGCTGCTCAGATGTCAACAAGGCATTTGAGCAGTTTGTGAAGC 910
DB 431 TAACTTGGAGATAAAGCTGCTCAGATGTCAACAAGGCATTTGAGCAGTTTGTGAAGC 490
QY 911 CGGAACAGCTTGATGGAGAAACTCGTACAAGTCAGCAAGTCAGCAAGTGTAAAAGATGTTCC 968
DB 491 CGGAACAGCTTGATGGAGAAACTCGTACAAGTCAGCAAGTGTAAAAGATGTTCC 548

RESULT 30
BI914337
LOCUS
DEFINITION 603182326F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5246166 5',
mRNA sequence.
BI914337
VERSION
KEYWORDS
SOURCE EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 719)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1AM11621 row: d column: 07
High quality sequence stop: 535.
Location/Qualifiers
1. .719
/organism="Homo sapiens"

```

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5246166"
/lab_host="DH10B"
/clone_lib="NIH_MGC_121"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: EcoRV (destroyed); RNA source anonymous pool of 3
fetal brains, female age 20 weeks, female age 24 weeks,
and male age 26 weeks. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
0.7-3.5 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 017. Note:
this is a NIH_MGC Library."

```

ORIGIN

```

Query Match      18.0%; Score 433; DB 12; Length 719;
Best Local Similarity 100.0%; Pred. No. 3.2e-210;
Matches 433; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 60
DB 26 GGGCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 85
QY 61 CAGAGTTGAACAATGACCAATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTAT 120
DB 86 CAGAGTTGAACAATGACCAATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTAT 145
QY 121 CAGATCAGCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 180
DB 146 CAGATCAGCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 205
QY 181 AGCTGGGCTGTGTCTTCAATGATGTCGACCAATGTCGACCTGATTAATCAAAACCATCACA 240
DB 206 AGCTGGGCTGTGTCTTCAATGATGTCGACCAATGTCGACCTGATTAATCAAAACCATCACA 265
QY 241 GTACTGTGTGTAGTTTATTCGAGTTTCATCTGTACCTGATTAATCAAAACCATCACA 300
DB 266 GTACTGTGTGTAGTTTATTCGAGTTTCATCTGTACCTGATTAATCAAAACCATCACA 325
QY 301 CAAAGGATCAAGCCCTAGTGTGATGCGCTCTCTCCAGAGAGTTCTTTTCCCATCT 360
DB 326 CAAAGGATCAAGCCCTAGTGTGATGCGCTCTCTCCAGAGAGTTCTTTTCCCATCT 385
QY 361 GAGAGAGTTTGTCTTAAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTG 420
DB 386 GAGAGAGTTTGTCTTAAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTG 445
QY 421 GGCAATACCTGTT 433
DB 446 GGCAATACCTGTT 458

RESULT 31
BI418577/c
LOCUS
DEFINITION t169h12.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2104583 3',
mRNA sequence.
BI418577
VERSION BI418577.1 GI:4264508
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 466)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.

```



```

QY 2253 AGGCTCAGCGATGAAATAGTGCACCTGGAGCAGAGGGGCCCTCCGAGGACCGCGA 2312
Db 120 AGGCTCAGCGATGAAATAGTGCACCTGGAGCAGAGGGGCCCTCCGAGGACCGCGA 61
QY 2313 CGCGAGCCTCAGCTCGGAGCCCCCGCGCGAATCCTCGAGGAGCAGATCGGCGCG 2372
Db 60 CGCGAGCCTCAGCTCGGAGCCCCCGCGCGAATCCTCGAGGAGCAGATCGGCGCGC 1

RESULT 33
AG116030/c
LOCUS AG116030 657 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-123H03.R, genomic survey sequence.
ACCESSION AG116030
VERSION AG116030.1 GI:16736549
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library PTB
Unpublished
2 (bases 1 to 657)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpansegsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1. .657
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-123H03.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN
Query Match 17.2%; Score 415; DB 29; Length 657;
Best Local Similarity 99.5%; Pred. No. 5.4e-201;
Matches 565; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1457 AGAATCACCTCCTTAATGGGACTGGACCATTTGAAGACACGCCAAGCATGTCATGT 1516
Db 621 AGAATCACCTCCTTAATGGGACTGGACCATTTGAAGACACGCCAAGCATGTCATGT 562
QY 1517 CGAGTCTTAACGGGAATTCAGTGTCAACAGGGCTAGTCTCTGTTAAGCTTCAGCTTCTG 1576
Db 561 CGAGTCTTAACGGGAATTCAGTGTCAACAGGGCTAGTCTCTGTTAAGCTTCAGCTTCTG 502
QY 1577 TCCAAACTGGTCTAGTTAATAGGTCCTCAGTGAPCCAGAACATCCTTAAGAAAAA 1636
Db 501 TCCAAACTGGTCTAGTTAATAGGTCCTCAGTGATCCAGAACATCCTTAAGAAAAA 442
QY 1637 TTCAATCAGTATTCACAAAGTTGCTGTTGCCAGTGTCAAGCTTCAACCTTCAACCTTC 1696
Db 441 TTCAATCAGTATTCACAAAGTTGCTGTTGCCAGTGTCAAGCTTCAACCTTCAACCTTC 382

```

```

QY 1697 ATAGTAATTTCTTTGGAGAACCTTACCAAGCCGCTTCCCTCTTTACCAATTACCAATTCGT 1756
Db 381 ATAGCAATTTCTTTGGAGAACCTTACCAAGCCGCTTCCCTCTTTACCAATTACCAATTCGT 322
QY 1757 CAGTACAGTCTACCTCGAAGCGCATCTACGATGTCTAGTAAAGTAAACAAACCGA 1816
Db 321 CAGTACAGTCTACCTCGAAGCGCATCTACGATGTCTAGTAAAGTAAACAAACCGA 262
QY 1817 TCCCCCGCAGTGAATCCTGCTCCAGCCCGCTGATGAATGCAATCAAGTCAACTCCA 1876
Db 261 TCCCCCGCAGTGAATCCTGCTCCAGCCCGCTGATGAATGCAATCAAGTCAACTCCA 202
QY 1877 GCCTGTGTTGTCCTATGGCGCCGAGTCTCTGAGGACTCTGAGGAGTCTCAAGGGGC 1936
Db 201 GCCTGTGTTGTCCTATGGCGCCGAGTCTCTGAGGACTCTGAGGAGTCTCAAGGGGC 142
QY 1937 TGGGCAAGGAGATGGATTGGTACGATTGTGAGTCTCCCACTCTCCGGCCAAAGATGCCG 1996
Db 141 TGGGCAAGGAGATGGATTGGTACGATTGTGAGTCTCCCACTCTCCGGCCAAAGATGCCG 82
QY 1997 AAGATGAGGAGGCCACTTCCGCAAGAGCT 2024
Db 81 AAGATGAGGAGGCCACTTCCGCAAGAGCT 54

RESULT 34
AG121934
LOCUS AG121934 818 bp mRNA linear EST 01-AUG-2002
DEFINITION MAMMAL1 Homo sapiens cDNA clone MAMMAL1001308 5', mRNA
sequence.
ACCESSION AG121934
VERSION AG121934.1 GI:10937169
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saiko, K., Kawai, Y.,
Yanamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and
Isogai, T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1. .818
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MAMMAL1001308"
/tissue_type="mammary gland"
/clone_lib="MAMMAL1"
/note="Vector: pME18SFL3"

ORIGIN
Query Match 16.9%; Score 407; DB 9; Length 818;
Best Local Similarity 99.7%; Pred. No. 7.2e-197;
Matches 577; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 798 CAAATGTTTAAATGCAAGGCCGCTTTCAGATATCTTTGATCCATATCTTGATATAACATT 857
Db 72 CAAATGTTTAAATGCAAGGCCGCTTTCAGATATCTTTGATCCATATCTTGATATAACATT 131

```

/clone lib="NIH_MGC_71"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 2.1 kb."

ORIGIN	Query Match	16.7%	Score 404	DB 10	Length 822
	Best Local Similarity	100.0%	Pred. No. 2.5e-195		
	Matches 404	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	933	CTCGTACAAAGTCAGCAAGTGTAAAGAGATGCTCCAGCTTCAAGAGAGTTCACATATCCA	992		
Db	1	CTCGTACAAAGTCAGCAAGTGTAAAGAGATGCTCCAGCTTCAAGAGAGTTCACATATCCA	60		
QY	993	TAGATCTCTTAATGTTCTTACACTTCTCTGAAACGTTTTGCAAAATTTTACCGGTGAAA	1052		
Db	61	TAGATCTCTTAATGTTCTTACACTTCTCTGAAACGTTTTGCAAAATTTTACCGGTGAAA	120		
QY	1053	AATTGCTAAAGGATGTGAAATACCTGAGTATCTTGATATTCGGCCATATATGCTCAACC	1112		
Db	121	AATTGCTAAAGGATGTGAAATACCTGAGTATCTTGATATTCGGCCATATATGCTCAACC	180		
QY	1113	CAACGAGAGCCAAATGTTCTACGCTCTTGATGAGTCTGCTCCACACTGGTTTTAATTG	1172		
Db	181	CAACGAGAGCCAAATGTTCTACGCTCTTGATGAGTCTGCTCCACACTGGTTTTAATTG	240		
QY	1173	CCATGCTGGCCATTACTTCTGTACATATAAGCTAGCAATGGCTCTGTGATCAAAATGAA	1232		
Db	241	CCATGCTGGCCATTACTTCTGTACATATAAGCTAGCAATGGCTCTGTGATCAAAATGAA	300		
QY	1233	TGACTCCATTGATATACCAAGTGTATGATAGTCTGAGTCTAGCCCAAGCCATGCT	1292		
Db	301	TGACTCCATTGATATACCAAGTGTATGATAGTCTGAGTCTAGCCCAAGCCATGCT	360		
QY	1293	CTTTTATATCAGTCCCATGATGTGAAAAATGGAGGTGAACCTTA	1336		
Db	361	CTTTTATATCAGTCCCATGATGTGAAAAATGGAGGTGAACCTTA	404		

RESULT 36
AI024619/c
LOCUS
DEFINITION ov52901.x1 Soares_testis_NHT Homo sapiens cdna clone IMAGE:1640976
3', mRNA sequence.
ACCESSION AI024619
VERSION AI024619.1 GI:3240232
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 822)
AUTHORS NIH-MGC http://mgi.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Arrayed by: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM9718 row: j column: 24
High quality sequence stop: 643.
Location/Qualifiers
1. .822
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3907703"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"

FEATURES
source
1. .374
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1640976"

RESULT 35
BE884143
LOCUS
DEFINITION 601506093f1 NIH_MGC_71 Homo sapiens cdna clone IMAGE:3907703 5',
mRNA sequence.
ACCESSION BE884143
VERSION BE884143.1 GI:10332919
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 822)
AUTHORS NIH-MGC http://mgi.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Arrayed by: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM9718 row: j column: 24
High quality sequence stop: 643.
Location/Qualifiers
1. .822
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3907703"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"

```

/sex="male"
/lab_host="DH10B"
/clone_lib="Soares testis NHT"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo (dT)
primer [5].
TCTTACCAATCTGAAGTGGAGCGCGGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match      15.5%; Score 374; DB 9; Length 374;
Best Local Similarity 100.0%; Pred. No. 5.2e-180;
Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1999 GATGAGGAGGCCACTCCGACGAGCTTCAGAACCCATGACCTTAACCGTCTAATAGT 2058
Db      |||||
QY 374 GATGAGGAGGCCACTCCGACGAGCTTCAGAACCCATGACCTTAACCGTCTAATAGT 315
Db      |||||
QY 2059 GCAGACAGCGACAGTACCCGAGAAACGGCTAGCGCTGATGGTCCAGCTGCCAA 2118
Db      |||||
QY 314 GCAGACAGCGACAGTACCCGAGAAACGGCTAGCGCTGATGGTCCAGCTGCCAA 255
Db      |||||
QY 2119 GCGAGCTGCGCTGCACTCAGAAATCCCTTTGCTAAGGCAACGGTCTTCTCGAAAG 2178
Db      |||||
QY 254 GCGCAGCGCTGCGCTGCACTCAGAAATCCCTTTGCTAAGGCAACGGTCTTCTCGAAAG 195
Db      |||||
QY 2179 TTGATGCTGCTCTTGTCTGCTCTCCAGAGACAAATCTTAGACCTTACGGCTT 2238
Db      |||||
QY 194 TTGATGCTGCTCTTGTCTGCTCTCCAGAGACAAATCTTAGACCTTACGGCTT 135
Db      |||||
QY 2239 AGCAACAACTGAAAGGCTCGACGATGAATGAGTGCACCTGGAGAGAGGGGCCCT 2298
Db      |||||
QY 134 AGCAACAACTGAAAGGCTCGACGATGAATGAGTGCACCTGGAGAGAGGGGCCCT 75
Db      |||||
QY 2299 CCGAGGACCGGAGCGCGAGCTCAGCTGCGACCCCGCGCGAATCCCTGGAGGAG 2358
Db      |||||
QY 74 CCGAGGACCGGAGCGCGAGCTCAGCTGCGACCCCGCGCGAATCCCTGGAGGAG 15
Db      |||||
QY 2359 CCAGATGGCGCGC 2372
Db      |||||
QY 14 CCAGATGGCGCGC 1
Db      |||||

RESULT 37
BF513259/c
LOCUS
DEFINITION
UT-H-BWL-amo-c-08-0-UI-sl NCI_CGAP_Sub7 Homo sapiens cDNA clone
IMAGE:3070478 3', mRNA sequence.
BF513259
VERSION
BF513259.1 GI:11598438
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 379)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: NCI-CGAP clone distribution
Information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

```

```

Seq primer: M13 Forward
POLYA=No.

FEATURES
source
Location/Qualifiers
1..379
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3070478"
/clone_lib="NCI_CGAP_Sub7"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; NCI_CGAP_Sub7
is a subtracted library derived from NCI_CGAP_Sub6. The
NCI_CGAP_Sub7 library had 12 million recombinants. A
single-stranded DNA preparation of NCI_CGAP_Sub6 was used
as a tracer in a subtractive hybridization with a driver
comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLAM
3334-3337, 3682-3683, 3798-3803 (IMAGE CloneIDs
1322376-1323911, 1456008-1456775, 1500552-1502855);
NCI_CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
(IMAGE CloneIDs 1323912-1325831, 1471368-1472903,
1492104-1493255); NCI_CGAP_Lu5 pool 1 LLAM 3575-3582,
3851-3854 (IMAGE CloneIDs 1414920-1417991,
1520904-1522439); NCI_CGAP_GC4 pool 1 LLAM 3164-3167,
3716-3720, 3733-3735 (IMAGE CloneIDs
1257096-1258631, 1469064-1470983, 1475592-1476743);
NCI_CGAP_P22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068
(IMAGE CloneIDs 985608-986759, 1101192-1101959,
1217928-1220615); NCI_CGAP_Co10 pool 1 LLAM 2644-2653,
2871-2872 (IMAGE CloneIDs
1057416-1061255, 1144584-1145351). (6% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI_CGAP_Sub1 (IMAGE CloneIDs 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE CloneIDs 2710536-2712455) (4% of
the driver population), plus a pool of 11,136 clones from
NCI_CGAP_Sub3 (IMAGE CloneIDs 2712456-2723591) (10% of
the driver population), plus a pool of 5,472 clones from
NCI_CGAP_Sub4 (IMAGE CloneIDs 2723592-2729326) (40% of the
driver population), plus a pool of 4032 clones from
NCI_CGAP_Sub6 (IMAGE CloneIDs 2728969-2733190) (40% of the
driver population). Subtraction was performed as
previously described [Bonaldo, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_SEQ=None found"

ORIGIN

Query Match      15.3%; Score 370; DB 10; Length 379;
Best Local Similarity 100.0%; Pred. No. 5.9e-178;
Matches 370; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2002 GAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCTTAACCGTCTAATAGTCA 2061
Db      |||||
QY 370 GAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCTTAACCGTCTAATAGTCA 311
Db      |||||
QY 2062 GACAGCGACAGTGACCCGAGAAACGGCTAGCGCTGATGTCGAGCTGCCAAGGC 2121
Db      |||||
QY 310 GACAGCGACAGTGACCCGAGAAACGGCTAGCGCTGATGTCGAGCTGCCAAGGC 251
Db      |||||
QY 2122 CAGCCTGCCCTGCACTCAGAAATCCCTTTGCTAAGCGAAACGGTCTTCTGGAAAGTTG 2181
Db      |||||
QY 250 CAGCCTGCCCTGCACTCAGAAATCCCTTTGCTAAGCGAAACGGTCTTCTGGAAAGTTG 191
Db      |||||
QY 2182 ATGCTCTCTCTTTGCTGCTCTCTCCAGAGACAAATCTTAGAGACCTTCAGGCTTAGC 2241
Db      |||||
QY 190 ATGCTCTCTCTTTGCTGCTCTCTCCAGAGACAAATCTTAGAGACCTTCAGGCTTAGC 131
Db      |||||
QY 2242 AACAACTGAAAGGCTCGACGAGTGAATGATGTCACCTGGAGCAGAGAGGGGCCCTCCC 2301
Db      |||||
QY 130 AACAACTGAAAGGCTCGACGAGTGAATGATGTCACCTGGAGCAGAGAGGGGCCCTCCC 71
Db      |||||
QY 2302 GAGGACCGGACCGCGAGCTCAGCTGGCAGCCCGCGCGAATCCCTGGAGGAGCCA 2361
Db      |||||

```

Db	70	GAGGACCGCAGCGAGCCTCAGCCTGGCAGCCCGCGCGGAGTCCCTGGAGGAGCCA	11
Qy	2362	GATGGCGCC 2371	
Db	10	GATGGCGCC 1	
RESULT 38			
LOCUS	BF063690/c		
DEFINITION	7h98c12.x1 NCI_CGAP_Col16 Homo sapiens cDNA clone IMAGE:3324022 3', mRNA sequence.	453 bp	mRNA linear EST 16-OCT-2000
ACCESSION	BF063690		
VERSION	BF063690.1	GI:10822600	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 453)		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgabbs@mail.nih.gov Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov Seq primer: -40UP from Gibco High quality sequence stop: 402.		
FEATURES	source	1..453	Location/Qualifiers
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:3324022"	
		/tissue_type="colon tumor, RER+"	
		/lab_host="DH10B"	
		/clone_lib="NCI_CGAP Col16"	
		/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP Col10 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonedDs 1057416-1061255, and 1144584-1145351). Subtraction by Bento Soares and M. Fatima Bonaldo."	
ORIGIN			
	Query Match	15.2%;	Score 366; DB 10; Length 453;
	Best Local Similarity	99.8%;	Pred. No. 7e-176;
	Matches	416;	Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy	1920	CGAGAGTCAAGGGCTGGCGCAGGAGAAATGGGATTTGGTACGATTGTGAGCTCCCACTC	1979
Db	453	CGAGAGTCAAGGGCTGGCGCAGGAGAAATGGGATTTGGTACGATTGTGAGCTCCCACTC	394
Qy	1980	TCCCGGCCAAGATGCCGAAGATGAGAGGCCCACTCCGACAGCTTCAAGAACCCATGAC	2039
Db	393	TCCCGGCCAAGATGCCGAAGATGAGAGGCCCACTCCGACAGCTTCAAGAACCCATGAC	334
Qy	2040	CCTAAACGGTGTATAGTGTGAGCAGCAGCAGTACCCGAAAGAACGGCCTAGCGCC	2099
Db	333	CCTAAACGGTGTATAGTGTGAGCAGCAGCAGTACCCGAAAGAACGGCCTAGCGCC	274
Qy	2100	TGATGGTCCAGCTGCCAAGGCGCAGCTGCCCTGCACTCAGAAAATCCCTTTGCTAAGGC	2159
Db	273	TGATGGTCCAGCTGCCAAGGCGCAGCTGCCCTGCACTCAGAAAATCCCTTTGCTAAGGC	214
Qy	2160	AAACGGTCTTCTGGAAGTTGATCGCTCTCTTTGCTGTCTCTCCCAAGACAAAAAT	2219
Db	213	AAACGGTCTTCTGGAAGTTGATCGCTCTCTTTGCTGTCTCTCCCAAGACAAAAAT	154
Qy	2220	CTTAGAGACCTTCAGGCTTAGCAACAACTGAAAGGCTCGACGGATGAAATGAGTGCACC	2279
Db	153	CTTAGAGACCTTCAGGCTTAGCAACAACTGAAAGGCTCGACGGATGAAATGAGTGCACC	94
Qy	2280	TGGAGCAGAGAGGGGCTCTCCGAGGACCGGACGCGGAGCTCAGCTTGGCAGCCC	2336
Db	93	TGGAGCAGAGAGGGGCTCTCCGAGGACCGGACGCGGAGCTCAGCTTGGCAGCCC	37
RESULT 39			
LOCUS	AA255809	416 bp	mRNA linear EST 06-AUG-1997
DEFINITION	5' mRNA sequence.		
ACCESSION	AA255809		
VERSION	AA255809.1	GI:1891350	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 416) Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lemmon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.		
TITLE	WashU-Merck EST Project 1997		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 849 Std Error: 0.00 Seq primer: -28m13 rev2 ET from Amersham.		
FEATURES	source	1..416	Location/Qualifiers
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:682343"	
		/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"	
		/lab_host="DH10B"	
		/clone_lib="Soares NHMPU_S1"	
		/note="Organ: mixed (see below); Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbHM, pregnant uterus NBHPU, and fetal heart NBH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."	
ORIGIN			
	Query Match	14.8%;	Score 358; DB 9; Length 416;
	Best Local Similarity	100.0%;	Pred. No. 8.7e-172;
	Matches	358;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1538	GTGTCAACAGGCTGTGCTCTTAATGCTTCAGCTTCAGTCCAAAACCTGGTCAGTTAATA	1597

/notes="Organ: placenta; Vector: pluscriptPR; Site: 1: all-XhoI; Site 2: BamH; Oligo-d primed using primer 5'-TTTTTTTTTTTTTTTT-3', size-selected for average insert size 2.3 kb and normalised to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: This is a NIH MGCI library."

ORIGIN	Query Match	14.1%;	Score 340;	DB 14;	Length 861;
	Best Local Similarity 99.7%;	Pred. No. 1.8e-162;	Indels 0;	Gaps 0;	
	Matches 390;	Conservative 0;	Mismatches 1;		
Qy	63	GAGTGTGAACAATGACCATAGTGTGACAAAGCTTCTGAAATCTTCAGACCCATCAGCCTATCA	122		
Db	80	GAGTGTGAACAATGACCATAGTGTGACAAAGCTTCTGAAATCTTCAGACCCATCAGCCTATCA	139		
Qy	123	GAATCAGCCTGGCAGCTCCGAGGCAGTCTCACCTGGAGAGCATGATGCAGGTTCTGCGCAG	182		
Db	140	GAATCAGCCTGGCAGCTCCGAGGCAGTCTCACCTGGAGAGCATGATGCAGGTTCTGCGCAG	199		
Qy	183	CTGGGTGCTGTGTCTTCATTGAATGATGTGCTCAAAATCACACATCTTTCTTAGGACCAAGT	242		
Db	200	CTGGGTGCTGTGTCTTCATTGAATGATGTGCTCAAAATCACACATCTTTCTTAGGACCAAGT	259		
Qy	243	ACCTGTGTGCTGTAGTCTTATTCGAGTTCTCATCTGTACCTGTATAATCAAAACCATCACCACA	302		
Db	260	ACCTGTGTGCTGTAGTCTTATTCGAGTTCTCATCTGTACCTGTATAATCAAAACCATCACCACA	319		
Qy	303	AAAGGATCAAGCCCTAGTGATGGCATCGCTCCTCCACAGAAAGTCTTTTCCCATCTGA	362		
Db	320	AAAGGATCAAGCCCTAGTGATGGCATCGCTCCTCCACAGAAAGTCTTTTCCCATCTGA	379		
Qy	363	GAAGATTTGCTTAAAGTGGCAACAACCTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGG	422		
Db	380	GAAGATTTGCTTAAAGTGGCAACAACCTCATAAAGTTGGAGCTGGGCTCCAGAAATTTGGG	439		
Qy	423	CAATACCTGTTTTCGAATGAGCACTGCGAG	453		
Db	440	CAATACCTGTTTTCGAATGAGCACTGCGAG	470		

RESULT 42
BX331803
LOCUS
BX331803 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
DEFINITION BX331803 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
linear mRNA
1144 bp
blast -ccncccacgagcctgggga 5-ppptttt
mRNA sequence.

ACCESSION BX331803
 VERSION BX331803.1 GI:30339041
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1144)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 CONTACT Genoscope
 Genoscope -- Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: secref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 4596.f For
 more information about this cluster, see
[http://www.genoscope.cns.fr/](http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DC002D010QPLcluster=4596.f)
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DC002D010QPLcluster=4596.f>. Contact :
 Feng Liang Email : liang@lifetech.com URL : <http://www.invitrogen.com>
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600
 Paraday Avenue Genoscope sequence ID : CS0DC002D010QPL.
 Location/Qualifiers

FEATURES

1. 1144
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSDDC0027H20"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="first strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized." 1

[illegible]

RESULT 43	AA016255	499 bp	mRNA	linear	EST 29-NOV-1996
LOCUS	AA016255/c	ze33c07.s1	Soares retina N2b4HR	homo sapiens	CDNA clone
DEFINITION	IMAGE:360780	3', mRNA sequence.			
ACCESSION	AA016255				
VERSION	AA016255.1	GI:1477302			

EST. 1952-1953

KEYWORDS	
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 499)
AUTHORS	Hillier, L., Clark, N., Dubuque, T., Eliston, K., Hawkins, M., Folman, M., Hultman, M., Kucaba, T., Le, M., Lemon, G., Marra, M., Parsons, J., Rifkin, L., Rohlif, T., Soares, M., Tan, F., Treviski, B., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE	COMMENT
<p>WILSON, K. The WashU-Merck EST Project Unpublished (1995) Contact: Wilson RK Washington University School 4444 Forest Park Parkway, B</p>	

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800			
Fax: 314 286 1810			
Email: est@watson.wustl.edu			
This clone is available royalty-free through LINL ; contact the			
IMAGE Consortium (info@image.lnl.gov) for further information.			
Possible reversed clone: polyt not found			
Insert Length: 1460 Std Error: 0.00			
Seq primer: -40M13 fwd. from Amersham			
High quality sequence stop: 347.			
Location/Qualifiers			
1. 499			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="GDB:1277325"			
/db_xref="taxon:9606"			
/clone="IMAGE:360780"			
/sex="male"			
/tissue_type="retina"			
/dev_stage="55 year old"			
/lab_host="DH10B (ampicillin resistant)"			
/clone_lib="Soares retina N2b4HR"			
/note="Organ: eye; Vector: pT73D (Pharmacia) with a			
modified polylinker; Site_1: Not 1; Site 2: Eco RI; 1st			
strand cDNA was primed with a Not I - oligo(dT) primer [5'			
TGTTACCAATCTGAAGTGGAGCGCGCGCTTTTTTTTTTTTTTTT 3'],			
double-stranded cDNA was size selected, ligated to Eco RI			
adapters (Pharmacia), digested with Not I and cloned into			
the Not I and Eco RI sites of a modified pT73 vector			
(Pharmacia). The retinas were obtained from a 55 year old			
Caucasian and total cellular poly(A)+ RNA was extracted 6			
hrs after their removal. The retina RNA was kindly			
provided by Roderick R. McInnes M.D. Ph.D. from the			
University of Toronto. Library constructed by Bento			
Soares and M.Fatima Bonaldo. "			
ORIGIN			
Query Match 13.1%; Score 317; DB 9; Length 499;			
Best Local Similarity 100.0%; Pred. No. 1e-150;			
Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	2052	TAATAGTGCACAGCGACAGTACCAGAAAGAAAACGGCTAGCGCTGATGGTCCAG	2111
Db	321	TAATAGTGCACAGCGACAGTACCAGAAAGAAAACGGCTAGCGCTGATGGTCCAG	262
QY	2112	CTGCCAAGGCCACCTGCCTCTGCTCTCTCCAGAAAGCAAAATCCTTGTAGCGAAAACGGTCTTCC	2171
Db	261	CTGCCAAGGCCACCTGCCTCTGCTCTCTCCAGAAAGCAAAATCCTTGTAGCGAAAACGGTCTTCC	202
QY	2172	TGAAAGTTGAGCCCTGCTCTTGTCTCTCTCCAGAAAGCAAAATCCTTGTAGCGAAAACGGTCTTCC	2231
Db	201	TGAAAGTTGATGCTCTGCTCTTGTCTCTCTCCAGAAAGCAAAATCCTTGTAGCGAAAACGGTCTTCC	142
QY	2232	CAGGCTTAGCAACAACTGAAAGGCTCGACCGATGAATGAGTCACCTGGAGCAGAGAG	2291
Db	141	CAGGCTTAGCAACAACTGAAAGGCTCGACCGATGAATGAGTCACCTGGAGCAGAGAG	82
QY	2292	GGGCGCTCCCGAGACCGCGACCGAGCTCAGCTGCGACGCCGCCGCGCAATCCCT	2351
Db	81	GGGCGCTCCCGAGACCGCGACCGAGCTCAGCTGCGACGCCGCCGCGCAATCCCT	22
QY	2352	GGAGGAGCCAGATGCGG	2368
Db	21	GGAGGAGCCAGATGCGG	5
RESULT 44			
AA456513			
LOCUS			
DEFINITION			
aai3d07.r1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:813133			
5', mRNA sequence.			
ACCESSION			
AA456513			
VERSION			
AA456513.1 GI:2179089			
KEYWORDS			
EST.			

LOCUS	AA016989	451 bp	mRNA	linear	EST 29-NOV-1999
DEFINITION	ze33c07.r1 Soares retina N2b4HR Homo sapiens cDNA clone IMAGE:360780 5', mRNA sequence.				

[illegible]

SOURCE: Homo sapiens

ORGANISM: Homo sapiens
Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eumycota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 49.)

AUTHORS: Hillier, L., Clark, N., Dubucq, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lemon, G., Maria, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.

ACCESSION	BU429862	
VERSION	BU429862.1	GI:22768349
KEYWORDS	EST.	
SOURCE	Homo sapiens	(human)
ORGANISM	Homo sapiens	
	Eukaryota;	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Catarrhini; Hominoidea; Hominidae; Homo.	

TITLE The WashU-Merck Est Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

TITLE	JOURNAL	COMMENT
National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)	
Contact: Robert Strausberg, Ph.D.		
Email: cgabs-r@mail.nih.gov		
Eco RI site shown at the beginning of the sequence.		
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.		

Fax: 314 200 1888
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1460 Std Error: 0.00
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 437.

cDNA Library Arrayed by: M.B. Soares Lab
 DNA Sequencing by: M.B. Soares Lab
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: M13 Forward.

FEATURES	Location/Qualifiers
source	1. .491

FEATURES	Location/Qualifiers
source	1. .385

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="CDB:1277325"
/db_xref="taxon:9606"
/clone="IMAGE:360780"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares retina N2b4HR"
/note="Crgan: eye; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5 TGTACCAATCATGTAGTGGAGCGCGCGCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)+ RNA was extracted 4 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Fatima Bonaldo."

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:306811"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_lines="MGC85"
/lab_host="DH10B (LTI)"
/clone_lib="NH MGC 50"
/note="Vector: pVT3-Pac; Site 1: NotI; Site 2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(3.5-4.4kb). Directionally cloned. Cells provided by
Louis M. Staudt, Ph.D. Library preparation by Maria de
Nazaria Rinaldo, Ph.D. and M. Bento Soares, Ph.D."

```

ORIGIN

Query Match 11.6%; Score 279; DB 9; Length 491;
Percentile Similarity 100.0%; Pred. No. 3,2e-131;

QY	QY	QY	QY
155	CTGGACATGATGCAGGTTCTGCCAGTGGGGTCTGTCTTCATTGAATGATGTCT	214	
68	CTGGACATGATGCAGGTTCTGCCAGTGGGGTCTGTCTTCATTGAATGATGTCT	127	

	Matches	Local Similarity	Conservative	Mismatches	Indels	Gaps
Qy	138	CTCCGAGGCAGTCTACCTGGAGACATGATGACGAGTTCGCCAGCTGGGGTCTGTGTC	100.00%	0	0	0
Db	138	CTCCGAGGCAGTCTACCTGGAGACATGATGACGAGTTCGCCAGCTGGGGTCTGTGTC	100.00%	0	0	0
Qy	198	TTCAATTGAATGATGCTGCAAAATCACACACTTTCTTTAGGACACGATACCTGGTGTCTAGT	100.00%	0	0	0
Db	198	TTCAATTGAATGATGCTGCAAAATCACACACTTTCTTTAGGACACGATACCTGGTGTCTAGT	100.00%	0	0	0
Qy	258	TTATTCGAGTTCACTCTGTACTCTGATAAATCAAACACCATCAACCAAAAGGATCAAGCCCT	100.00%	0	0	0
Db	258	TTATTCGAGTTCACTCTGTACTCTGATAAATCAAACACCATCAACCAAAAGGATCAAGCCCT	100.00%	0	0	0
Qy	318	AGGTGATGGCATGCTCCTCCACAGAAAGTCTTTTCCCATCTGAGNAGAGATTTGTCTTAA	100.00%	0	0	0

215	CAAATCACACACTTCTTTAGNACCACTGGTCTGTAGTTTATTTCGAGTTTCATCTG	274
QY		
128	CAAATCACACACTTCTTTAGGACCACTGCTGGTCTGTAGTTTATTTCGAGTTTCATCTG	187
Db		
275	TACCTGATTAATCAAAACCATCACCAAAAAGGATCAAGCCCTAGTGTATGGCATCGCTC	334
QY		
188	TACCTGATTAATCAAAACCATCACCAAAAAGGATCAAGCCCTAGTGTATGGCATCGCTC	247
Db		
335	CTCCACAGAAAGTTCTTTTCCCATCTGAGAAGATTTGTCTTAAGTGGCAACAACTCATA	394
QY		
248	CTCCACAGAAAGTTCTTTTCCCATCTGAGAAGATTTGTCTTAAGTGGCAACAACTCATA	307
Db		

RESULT 46

```

Db      318 AGGTGATGGCATCCTCTCCACGAAAGTTCTTTTCCCATCTGAGAAGATTGTCTTAA 377
QY      378 GTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAA 416
Db      378 GTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAA 416

RESULT 47
W16759
LOCUS   zbl16f04.r1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
DEFINITION IMAGE:302239 5' similar to contains element MSR1 repetitive element
// mRNA sequence.
W16759
ACCESSION W16759.1 GI:1291140
VERSION   1
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE 1 (bases 1 to 517)
AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE     The WashU-Merck EST Project
JOURNAL   Unpublished (1995)
COMMENT   Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through L1NL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: EFPrimer
High quality sequence stop: 348.
FEATURES
    source
        1..517
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="GDB:1247171"
            /db_xref="taxon:9606"
            /clone="IMAGE:302239"
            /dev_stage="19 weeks"
            /lab_host="DH10B (ampicillin resistant)"
            /clone_lib="Soares fetal lung NbHL19W"
            /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAATGGAGCGCGCAATTTTTTTTTTTTTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same focus as the fetal heart library, Soares
fetal heart NBHL19W."

ORIGIN
Query Match 11.4%; Score 274; DB 14; Length 517;
Best Local Similarity 99.7%; Pred. No. 1.2e-128;
Matches 394; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY      1159 ACTGGTTTAAATGCCAGCTGGCCATTACTCTGCTACATAAAAGCTAGCAATGGCCTC 1218
Db      1 ACTGGTTTAAATGCCAGCTGGCCATTACTCTGCTACATAAAAGCTAGCAATGGCCTC 60
QY      1219 TGGTATCAATGAATGATCTCATTTGATATCCAGTGATATTAGATCGGTACTCAGCCAA 1278

```

```

Db      61 TGGTATCAATGAATGATCTCATTTGATATCTACCAAGTGATATTAGATCGGTACTCAGCCAA 120
QY      1279 CAAGCCTATGTGCTCTCTTTTATATATCAGGTCCCATCATGTGAAAAATGGAGGTGAACCTTACT 1338
Db      121 CAAGCCTATGTGCTCTCTTTTATATCAGGTCCCATCATGTGAAAAATGGAGGTGAACCTTACT 180
QY      1339 CATCCGACCCATAGCCCGCGCCAGTCTCTCTCCCGCCCGCTCATCTCAGTCAGCGGGTTGTC 1398
Db      181 CATCCGACCCATAGCCCGCGCCAGTCTCTCTCCCGCCCGCTCATCTCAGTCAGCGGGTTGTC 240
QY      1399 ACCAAACAAACAGGCTGCGCCAGGCTTTATCGGACCACAGCTTCCTCTCATGATAAAG 1458
Db      241 ACCAAACAAACAGGCTGCG-CAGGCTTTATCGGACCACAGCTTCCTCTCATGATAAAG 299
QY      1459 AATCCACCTCAGTTAAATGGGACTGGACCATTTGAAGACACGACGACGCTTCATCTGCG 1518
Db      300 AATCCACCTCAGTTAAATGGGACTGGACCATTTGAAGACACGACGACGCTTCATCTGCG 359
QY      1519 AGTCCTAAACGGGAATTCAGTGTCACACAGGGCTAG 1553
Db      360 AGTCCTAAACGGGAATTCAGTGTCACACAGGGCTAG 394

RESULT 48
W16759
LOCUS   AA749124/c
DEFINITION AA749124
ACCESSION AA749124.1 GI:27899082
VERSION   1
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE 1 (bases 1 to 389)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 875 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amerham
High quality sequence stop: 370.
FEATURES
    source
        1..389
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:1270688"
            /tissue_type="germinal center B cell"
            /lab_host="DH10B"
            /clone_lib="NCI CGAP GCBI"
            /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAATGGAGCGCGCTCATTTTTTTTTTTTTT-3',
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I

```

and Eco RI sites of the modified p773 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

```

ORIGIN
Query Match      11.1%; Score 267; DB 9; Length 389;
Best Local Similarity 99.7%; Pred. No. 4.4e-125;
Matches 387; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1943 AGGAGATGGGATGGTACGATTGTGAGCTCCACATCTCCCGGCCAAGATGCCGAGATG 2002
DB 389 AGGAGATGGGATGGTACGATTGTGAGCTCCACATCTCCCGGCCAAGATGCCGAGATG 330
QY 2003 AGGAGGCCACTCCGACGAGCTTCAGAACCCCATGACCCCTAAACGGTGTATAGTGCAG 2062
DB 329 AGGAGGCCACTCCGACGAGCTTCAGAACCCCATGACCCCTAAACGGTGTATAGTGCAG 270
QY 2063 ACAGGACATGACCCGAGAACGAGGCTAGCGCTGATGTG-CCAGCTGCCAAGGC 2121
DB 269 ACAGGACATGACCCGAGAACGAGGCTAGCGCTGATGTGCGCCAGCTGCCAAGGC 210
QY 2122 CAGCGTCCCTGACCTCAGAAAATCCCTTTGTAAAGCAAAACGGTCTTCTCTGAAAGTTG 2181
DB 209 CAGCGTCCCTGACCTCAGAAAATCCCTTTGTAAAGCAAAACGGTCTTCTCTGAAAGTTG 150
QY 2182 ATGCCTGCTCCTTTGCTCTCTCCAGAGACAAATCTTAGAGACCTTCAGGCTTAGC 2241
DB 149 ATGCCTGCTCCTTTGCTCTCTCCAGAGACAAATCTTAGAGACCTTCAGGCTTAGC 90
QY 2242 AACAACTGAAAGCTGACGGATGAATGAGTGCACCTGAGCAGAGAGGGGCGCTCCC 2301
DB 89 AACAACTGAAAGCTGACGGATGAATGAGTGCACCTGAGCAGAGAGGGGCGCTCCC 30
QY 2302 GAGGACCGCAGCGCGGACCTTCAGCCTG 2329
DB 29 GAGGACCGCAGCGCGGACCTTCAGCCTG 2

```

```

RESULT 49
W19122      476 bp      mRNA      linear      EST 03-MAY-1996
LOCUS      zbl5b04.r1 Soares fetal lung NbHL19W Homo sapiens cDNA clone
DEFINITION IMAGE:302095 5' similar to contains element MSRI repetitive element
            i, mRNA sequence.
ACCESSION  W19122
VERSION     W19122.1 GI:1294865
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 476)
            Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
            Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
            Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
            Trevasakis, E., Waterston, R., Williamson, A., Wohldmann, P. and
            Wilson, R.
            The WashU-Merck EST Project
            Unpublished (1995)
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: ETPRimer
            High quality sequence stop: 373.
            Location/Qualifiers
FEATURES    1. .476
            /organism="Homo sapiens"
            /mol_type="mRNA"

```

```

/db_xref="GDB:1247027"
/db_xref="taxon:9606"
/clone="IMAGE:302095"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/clone_1_b="Soares fetal lung NbHL19W"
/note="Organ: lung; Vector: p773D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
15'-TGTTACCAATCTGAAGTGGAGCGCGCAATTTTTTTTTTTT-3',
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p773 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbHL19W."

```

```

ORIGIN
Query Match      10.9%; Score 264; DB 14; Length 476;
Best Local Similarity 99.7%; Pred. No. 1.6e-123;
Matches 384; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1159 ACTGGTTTAAATGGCCATGCTGGCCATTTACTTCTGTACATAAAAGCTAGCAATGGCCTC 1218
DB 1 ACTGGTTTAAATGGCCATGCTGGCCATTTACTTCTGTACATAAAAGCTAGCAATGGCCTC 60
QY 1219 TGTATCAATGAATGACTCCATTTATATCAGTCCCATGATGTGAAAATGAGGTGAATCTTACT 1278
DB 61 TGTATCAATGAATGACTCCATTTATATCAGTCCCATGATGTGAAAATGAGGTGAATCTTACT 180
QY 1279 CAAGCTTATGTCTCTTTTATATCAGTCCCATGATGTGAAAATGAGGTGAATCTTACT 1338
DB 121 CAAGCTTATGTCTCTTTTATATCAGTCCCATGATGTGAAAATGAGGTGAATCTTACT 180
QY 1339 CATCCCAACCATAGCCCCGGCCAGTCTCTCCCGCCCGCTCATCATGTCAGCGGGTGTG 1398
DB 181 CATCCCAACCATAGCCCCGGCCAGTCTCTCCCGCCCGCTCATCATGTCAGCGGGTGTG 240
QY 1399 ACACAAACAGCTGCGCCAGGCTTTATCGGACACAGCTTCCTCTCACATGATAAG 1458
DB 241 ACCAAACAAACAGCTGCGC-CAGGCTTTATCGGACACAGCTTCCTCTCACATGATAAG 299
QY 1459 AATCCACTCTCACTTAAATGGGACTTGGACCATTTGAAGACACGCCCAAGCAGTTCATGTCG 1518
DB 300 AATCCACTCTCACTTAAATGGGACTTGGACCATTTGAAGACACGCCCAAGCAGTTCATGTCG 359
QY 1519 AGTCCTAAACGGAAATTCAGTGTCA 1543
DB 360 AGTCCTAAACGGAAATTCAGTGTCA 384

```

```

RESULT 50
W16787      390 bp      mRNA      linear      EST 29-APR-1996
LOCUS      zbl6g05.r1 Soares fetal lung NbHL19W Homo sapiens cDNA clone
DEFINITION IMAGE:302264 5' similar to contains element MSRI repetitive element
            i, mRNA sequence.
ACCESSION  W16787
VERSION     W16787.1 GI:1291168
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 390)
            Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
            Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
            Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
            Trevasakis, E., Waterston, R., Williamson, A., Wohldmann, P. and
            Wilson, R.
            The WashU-Merck EST Project

```

JOURNAL
COMMENT

Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: ETPRimer
High quality sequence stop: 312.

FEATURES

source
1. .390
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:1247196"
/db_xref="taxon:9606"
/clone="IMAGE:302284"
/dev_stage="19 weeks"
/lab_host="PH10B (ampicillin resistant)"
/clone_lib="Soares fetal lung NbHL19W"
/note="Organ: lung; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
15'-TGTTACCAATCTGAAGTCGAGCGCGCAATTTTCTTTT-3',
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbHL19W."

ORIGIN

Query Match 10.7%; Score 259; DB 14; Length 390;
Best Local Similarity 100.0%; Pred. No. 5.6e-121;
Matches 259; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1159 ACTGGTTTAAATGGCATGCTGGCCATTACTTCTGCTACATAAAAGCTAGCAATGGCCTC 1218
Db 1 ACTGGTTTAAATGGCATGCTGGCCATTACTTCTGCTACATAAAAGCTAGCAATGGCCTC 60
Qy 1219 TGGTATCAAAATGAATGACTCCATGTTGATCTACAGTGATATTAGATCGGTACTCAGCCAA 1278
Db 61 TGGTATCAAAATGAATGACTCCATGTTGATCTACAGTGATATTAGATCGGTACTCAGCCAA 120
Qy 1279 CAAGCCTATGTGCTCTTTTATATCAGGTCCCATGATGTGAAAAATGGAGGTCAACTTACT 1338
Db 121 CAAGCCTATGTGCTCTTTTATATCAGGTCCCATGATGTGAAAAATGGAGGTCAACTTACT 180
Qy 1339 CATCCACCCATAGCCCGGCGAGTCTCTCCCGCCCGTCAATCAGTCAGGGGTGTGTC 1398
Db 181 CATCCACCCATAGCCCGGCGAGTCTCTCCCGCCCGTCAATCAGTCAGGGGTGTGTC 240
Qy 1399 ACCAACAAACAGGTGCGC 1417
Db 241 ACCAACAAACAGGTGCGC 259

Search completed: August 17, 2004, 07:20:10
Job time : 9159 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 16, 2004, 23:48:34 ; Search time 198 Seconds
(without alignments)
6760.311 Million cell updates/sec

Title: US-10-049-745-31

Perfect score: 2412

Sequence: 1 gggctgtgtgctggcgcgcgccg.....agggttaatttttagcttgca 2412

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 30

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	357	14.8	722	4	US-09-280-116-223	Sequence 223, App
2	205	8.5	457	4	US-09-280-116-214	Sequence 214, App
3	30	1.2	847	4	US-09-495-050A-40	Sequence 40, Appl
4	30	1.2	1273	4	US-09-976-594-448	Sequence 448, App

ALIGNMENTS

RESULT 1

US-09-280-116-223

; Sequence 223, Application US/09280116A

; Patent No. 6331427

; GENERAL INFORMATION:

; APPLICANT: Robison, Keith E.

; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs

; FILE REFERENCE: 5800-24, 035800/176965

; CURRENT APPLICATION NUMBER: US/09/280.116A

; CURRENT FILING DATE: 1999-03-26

; NUMBER OF SEQ ID NOS: 268

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 223

; LENGTH: 722

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: ubiquitin carboxyl-terminal hydrolases family 2

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)..(722)

; OTHER INFORMATION: n = a, t, c or g

US-09-280-116-223

Query Match 14.8%; Score 357; DB 4; Length 722;

Best Local Similarity 100.0%; Pred. No. 9.7e-173;

Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1061 AGGATGTGAAATACCCCTGAGTATCTTGATATTCGGCCATATATGTCTCAACCCCAACGGAG 1120

Db 107 AGGATGTGAAATACCCCTGAGTATCTTGATATTCGGCCATATATGTCTCAACCCCAACGGAG 166

QY 1121 AGCCAAATTTCTACGTCCTTTGTATGTCAGTGTGCTGCCACTGGTTCACACTGGTTCCTTTAATTTGCCATGCTG 1180

Db 167 AGCCAAATTTCTACGTCCTTTGTATGTCAGTGTGCTGCCACTGGTTCACACTGGTTCCTTTAATTTGCCATGCTG 226

QY 1181 GCATTTACTTCTGCTACATAAAAGCTAGCAATGGCCCTCTGGTATCAAAATGAATGACTCCA 1240

Db 227 GCATTTACTTCTGCTACATAAAAGCTAGCAATGGCCCTCTGGTATCAAAATGAATGACTCCA 286

QY 1241 TTGTATCTACCACTGATATTAGATCGGTACTCAGCCCAACAAAGCCTATGTCTCTTTTATA 1300

Db 287 TTGTATCTACCACTGATATTAGATCGGTACTCAGCCCAACAAAGCCTATGTCTCTTTTATA 346

QY 1301 TCAGTTCCTCATGTGTAATAATGGAGGTGAATCTTACTCATCCACCATAGCCCGGCC 1360

Db 347 TCAGTTCCTCATGTGTAATAATGGAGGTGAATCTTACTCATCCACCATAGCCCGGCC 406

QY 1361 AGTCCTCTCCCGCCCGCCGTCATCAGTCAGCGGGTGTGCACCAACAAAGGTGCGC 1417

Db 407 AGTCCTCTCCCGCCCGCCGTCATCAGTCAGCGGGTGTGCACCAACAAAGGTGCGC 463

RESULT 2

US-09-280-116-214

; Sequence 214, Application US/09280116A

; Patent No. 6331427

; GENERAL INFORMATION:

; APPLICANT: Robison, Keith E.

; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs

; FILE REFERENCE: 5800-24, 035800/176965

; CURRENT APPLICATION NUMBER: US/09/280.116A

; CURRENT FILING DATE: 1999-03-26

; NUMBER OF SEQ ID NOS: 268

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 214

; LENGTH: 457

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: ubiquitin carboxyl-terminal hydrolases family 2

US-09-280-116-214

Query Match 8.5%; Score 205; DB 4; Length 457;

Best Local Similarity 100.0%; Pred. No. 7.3e-95;

Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 CTCGAGGCGAGTCTCACCTGGAGACATGGATGCGGTTCTGCCAGCTGGGGTGTGTGTC 197

Db 103 CTCGAGGCGAGTCTCACCTGGAGACATGGATGCGGTTCTGCCAGCTGGGGTGTGTGTC 162

QY 198 TTCATTGAATGATGTGTCAAATCACACATTTCTTTAGGACCACTACCTGCTGTGTAGT 257

Db 163 TTCATTGAATGATGTGTCAAATCACACATTTCTTTAGGACCACTACCTGCTGTGTAGT 222

QY 258 TTATTCAGTTCATCTGTACCTGATAAATCAAAACCATCACCAAAAGGATCAAGCCCT 317

Db 223 TTATTCAGTTCATCTGTACCTGATAAATCAAAACCATCACCAAAAGGATCAAGCCCT 282

QY 318 AGGTGATGGCATCGCTCTCCACAG 342

Wed Aug 18 13:52:59 2004

Db 283 AGGTGATGCGATCGCTCCTCCACAG 307

RESULT 3

US-09-495-050A-40/c
 ; Sequence 40, Application US/09495050A
 ; Patent No. 6492505
 ; GENERAL INFORMATION:
 ; APPLICANT: Roopa, Reddy
 ; APPLICANT: Guegler, Karl, J.
 ; APPLICANT: Au-Young, Janice
 ; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED P
 ; FILE REFERENCE: PA-0013 US
 ; CURRENT APPLICATION NUMBER: US/09/495,050A
 ; CURRENT FILING DATE: 2000-01-31
 ; PRIOR APPLICATION NUMBER: 60/118,318
 ; PRIOR FILING DATE: February 1, 1999
 ; NUMBER OF SEQ ID NOS: 305
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 40
 ; LENGTH: 847
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. 6492505 797777CTT
 US-09-495-050A-40

Query Match 1.2%; Score 30; DB 4; Length 847;
 Best Local Similarity 100.0%; Pred. No. 3.4e-05;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2383 CCCTTTAGTCAGGGTTAAATTTTAGCTTGCA 2412
 Db 37 CCCTTTAGTCAGGGTTAAATTTTAGCTTGCA 8

RESULT 4

US-09-976-594-448
 ; Sequence 448, Application US/09976594
 ; Patent No. 6673549
 ; GENERAL INFORMATION:
 ; APPLICANT: Furness, Michael
 ; APPLICANT: Buchbinder, Jenny
 ; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
 ; FILE REFERENCE: PA-0041 US
 ; CURRENT APPLICATION NUMBER: US/09/976,594
 ; CURRENT FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: 60/240,409
 ; PRIOR FILING DATE: 2000-10-12
 ; NUMBER OF SEQ ID NOS: 1143
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 448
 ; LENGTH: 1273
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. 6673549 522433CB1
 US-09-976-594-448

Query Match 1.2%; Score 30; DB 4; Length 1273;
 Best Local Similarity 100.0%; Pred. No. 3.5e-05;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2383 CCCTTTAGTCAGGGTTAAATTTTAGCTTGCA 2412
 Db 1238 CCCTTTAGTCAGGGTTAAATTTTAGCTTGCA 1267

Search completed: August 17, 2004, 07:23:31
 Job time : 201 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2004, 04:45:41 ; Search time 3503 Seconds
(without alignments)
3378.458 Million cell updates/sec

Title: US-10-049-745-31

Perfect score: 2412

Sequence: 1 gggctgtgcggcgccg.....agggttaatttagcttgca 2412

Scoring table:

Gapop_60.0 , Gapext 60.0

Searched: 3225727 seqs, 2453303834 residues

Word size : 30

Total number of hits satisfying chosen parameters: 34

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2321	96.2	4659	13	US-10-399-645-22
2	2287	94.8	2347	15	US-10-163-547-2
3	2249	93.2	3951	16	US-10-379-981-8
4	2210	91.6	4483	13	US-10-302-172-390
5	333	13.8	333	9	US-09-864-761-31553
6	314	13.0	498	9	US-09-864-761-15021
7	205	8.5	489	9	US-09-864-761-14092
8	198	8.2	198	9	US-09-864-761-30650
9	157	6.5	504	13	US-10-027-632-128282
10	157	6.5	504	16	US-10-027-632-128282
11	97	4.0	631	13	US-10-027-632-277717
12	97	4.0	631	16	US-10-027-632-277717
13	68	2.8	591	13	US-10-027-632-85072
14	68	2.8	591	16	US-10-027-632-85072

15	41	1.7	3981	16	US-10-379-981-11	Sequence 11, Appl
C 16	39	1.6	592	13	US-10-027-632-319601	Sequence 319601,
C 17	39	1.6	592	16	US-10-027-632-319601	Sequence 319601,
C 18	30	1.2	637	17	US-10-258-662-51	Sequence 51, Appl
19	30	1.2	733	15	US-10-149-819-39	Sequence 39, Appl
C 20	30	1.2	847	15	US-10-313-542-40	Sequence 40, Appl
21	30	1.2	972	9	US-09-349-015-17	Sequence 17, Appl
22	30	1.2	1112	15	US-10-219-664-13	Sequence 13, Appl
23	30	1.2	1112	9	US-09-880-192-16	Sequence 16, Appl
24	30	1.2	1112	15	US-10-427-348-16	Sequence 16, Appl
25	30	1.2	1273	9	US-09-981-353-187	Sequence 187, App
26	30	1.2	1273	10	US-09-919-039-191	Sequence 191, App
27	30	1.2	1273	15	US-10-133-013-59	Sequence 59, Appl
28	30	1.2	1702	9	US-09-349-015-2	Sequence 2, Appl
29	30	1.2	1702	15	US-10-219-664-2	Sequence 2, Appl
C 30	30	1.2	1833	9	US-09-880-192-32	Sequence 32, Appl
C 31	30	1.2	1833	15	US-10-427-348-32	Sequence 32, Appl
C 32	30	1.2	1826	10	US-09-971-392-216	Sequence 216, App
C 33	30	1.2	1987	9	US-09-818-143-20	Sequence 20, Appl
C 34	30	1.2	2040	15	US-10-198-846-10830	Sequence 10830, A

ALIGNMENTS

RESULT 1

US-10-399-645-22
; Sequence 22, Application US/10399645
; Publication No. US20040029249A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; LEE, Ernestine A.
; APPLICANT: HAFALIA, April J.A.; YUE, Henry
; APPLICANT: IAL, Preeti G.; YAO, Monique G.
; APPLICANT: LU, Yan; CHAWLA, Narinder K.
; APPLICANT: WARREN, Bridget A.; LU, Dyung Aina M.
; APPLICANT: BAUGHN, Mariah R.; DELEGEANE, Angelo M.
; APPLICANT: BURFORD, Neil; BOROWSKY, Mark L.
; APPLICANT: LEE, Sally; XU, Yuming
; APPLICANT: GRIFFIN, Jennifer A.; KALLICK, Deborah A.
; APPLICANT: GANDHI, Ameena R.; ARVIZU, Chandra S.
; APPLICANT: ISON, Craig H.; TANG, Y. Tom
; APPLICANT: AZIMZAI, Valda; ELLIOTT, Vicki S.
; APPLICANT: SWARNAKAR, Anita; RAMKUMAR, Jayalaxmi
; APPLICANT: NGUYEN, Damiel B.; TRIBOULEY, Catherine M.
; APPLICANT: LO, Terence P.; AU-YOUNG, Janice K.
; APPLICANT: THANGAVELU, Kavitha; KEARNEY, Liam
; TITLE OF INVENTION: PROTEASES
; FILE REFERENCE: PI-0263 USN
; CURRENT APPLICATION NUMBER: US/10/399,645
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: PCT/US01/51034
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: US 60/241,573
; PRIOR FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US 60/243,643
; PRIOR FILING DATE: 2000-10-25
; PRIOR APPLICATION NUMBER: US 60/245,256
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/248,395
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: US 60/249,826
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/252,303
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,981
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 4659
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

1042 ACCGTTGGAATAATGCTTAAAGATGTGAAATACCTCGAGTATCTTGATATTCGGCCATAT 1101
1030 ACCGTTGGAATAATGCTTAAAGATGTGAAATACCTCGAGTATCTTGATATTCGGCCATAT 1089
1102 ATGCTTCAACCAACGGAGAGCAATGCTTACGCTTGTATGCAAGTGTGCTGCTCCACACT 1161
1090 ATGCTTCAACCAACGGAGAGCAATGCTTACGCTTGTATGCAAGTGTGCTGCTCCACACT 1149
1162 GGTTTTAAATGTCATGCTGGCCATATCTTCTGTATACATAAAAGCTAGCAATGGCTCTGG 1221
1150 GGTTTTAAATGTCATGCTGGCCATATCTTCTGTATACATAAAAGCTAGCAATGGCTCTGG 1209
1222 TATCAAAATGAATGACTCCATGTATCTACAGTGTATATAGATCGGTACTCGACCAACAA 1281
1210 TATCAAAATGAATGACTCCATGTATCTACAGTGTATATAGATCGGTACTCGACCAACAA 1269
1282 GCTATGCTCTTTTATATCAGTCCCATGATGTGAAAAATGGAGGTGAATCTCAT 1341
1270 GCTATGCTCTTTTATATCAGTCCCATGATGTGAAAAATGGAGGTGAATCTCAT 1329
1342 CCCACCATAGCCCGCCAGTCTCTCCCGCCCGTCAATCAGTCAGCGGGTTGCACC 1401
1330 CCCACCATAGCCCGCCAGTCTCTCCCGCCCGTCAATCAGTCAGCGGGTTGCACC 1389
1402 AACAAACAGGCTGGCCAGGCTTTATCGACACAGTTCCTCTCACATGATAAAGAAAT 1461
1390 AACAAACAGGCTGGCCAGGCTTTATCGGACCAAGCTTCCCTCTCACATGATAAAGAAAT 1449
1462 CCACCTCACTTAATGGAGTGGACCATTGAAAGACAGCGCAAGCATTCATGTGAGT 1521
1450 CCACCTCACTTAATGGAGTGGACCATTGAAAGACAGCGCAAGCATTCATGTGAGT 1509
1522 CTTAAACGGGAATCCAGTGTCAACAGGCTAGTCTGTTAATGCTTCAGTCTCTGTCCAA 1581
1510 CTTAAACGGGAATCCAGTGTCAACAGGCTAGTCTGTTAATGCTTCAGTCTCTGTCCAA 1569
1582 AACTGTGTCAGTAAATAGGTCCTGATGATCCAGAAATCCAGAAATCCAGAAATTAACA 1641
1570 AACTGTGTCAGTAAATAGGTCCTGATGATCCAGAAATCCAGAAATCCAGAAATTAACA 1629
1642 ATCAGTATTCACACAAAGTTCCTGTTCCGAGGTGTCAGTCTCAACCTAACCTTCATAGT 1701
1630 ATCAGTATTCACACAAAGTTCCTGTTCCGAGGTGTCAGTCTCAACCTAACCTTCATAGT 1689
1702 AATTCCTTTGGAGAACCTTACCAAGCCCGTTCCTCTTCTACCATTAACAATTCGCGAGTA 1761
1690 AATTCCTTTGGAGAACCTTACCAAGCCCGTTCCTCTTCTACCATTAACAATTCGCGAGTA 1749
1762 CAGTCTACCTCGAAGCGCATCTACGATGTGAGTTCCTAGTAAAGTAAACAAACCGATCCCC 1821
1750 CAGTCTACCTCGAAGCGCATCTACGATGTGAGTTCCTAGTAAAGTAAACAAACCGATCCCC 1809
1822 CGCAGTGAATCCTGCTCCAGCCGCTGATGATGGAAGTAAATCCAAAGCTGAATCCAGCGTG 1881
1810 CGCAGTGAATCCTGCTCCAGCCGCTGATGATGGAAGTAAATCCAAAGCTGAATCCAGCGTG 1869
1882 CTGCTGCTTATGCGCCGAGTCTCTGAGGACTCTGAGGAGTCAAGGGGCTGGGC 1941
1870 CTGCTGCTTATGCGCCGAGTCTCTGAGGACTCTGAGGAGTCAAGGGGCTGGGC 1929
1942 AAGGAGATGGATGGTACGATGTTGAGTCCCACTCTCCGCGCAAGATGCCGAAGAT 2001
1930 AAGGAGATGGATGGTACGATGTTGAGTCCCACTCTCCGCGCAAGATGCCGAAGAT 1989
2002 GAGGAGGCGACCTCCGACAGCTTCAAGAACCCATGACCTTAAACCGTGCTTAATAGTGA 2061
1990 GAGGAGGCGACCTCCGACAGCTTCAAGAACCCATGACCTTAAACCGTGCTTAATAGTGA 2049
2062 GACAGCGACAGTGAACCCGAAAGAAACGGCCTAGGCTGATGCTGCGAGCTGCCAAGGC 2121
2050 GACAGCGACAGTGAACCCGAAAGAAACGGCCTAGGCTGATGCTGCGAGCTGCCAAGGC 2109
2122 CAGCCTGCGCTGCACTCAGAAATCCCTTTGTAAGGCAACCGGTCTTCTCGAAAGTTG 2181

2110 CAGCCTGCGCTGCACTCAGAAAAATCCCTTTGCTAAGGCAAAACGGTCTTCTCTGAAAAGTTG 2169
2182 ATGCTGCTCTCTTTGCTGCTCTCTCCGAAAGCAAAATCTTAGAGACCTTTCAGGCTTAGC 2241
2170 ATGCTGCTCTCTTTGCTGCTCTCTCCGAAAGCAAAATCTTAGAGACCTTTCAGGCTTAGC 2229
2242 AACAACTGAAAGGCTCGACGGATGAAATGATGCACTGGAGCAGAGAGGGGCCCTCCC 2301
2230 AACAACTGAAAGGCTCGACGGATGAAATGATGCACTGGAGCAGAGAGGGGCCCTCCC 2289
2302 GAGGACCGCAGCGCAGGCTCAGCCTGGCAGCCCGCGCGCGAATCCCTGCGAGGAGC 2359
2290 GAGGACCGCAGCGCAGGCTCAGCCTGGCAGCCCGCGCGCGAATCCCTGCGAGGAGC 2347
RESULT 3
US-10-379-981-8
; Sequence 8, Application US/10379981
; Publication No. US20040001820A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pharmaceuticals, Inc.
; APPLICANT: Hahn, Chang S
; APPLICANT: Liu, Hong S
; TITLE OF INVENTION: HUMAN DEUBIQUITINATING PROTEASE GENE ON CHROMOSOME 7 AND ITS
; TITLE OF INVENTION: MURINE ORTHOLOG
; FILE REFERENCE: USAV2002-0022 WO PCT
; CURRENT APPLICATION NUMBER: US/10/379,981
; CURRENT FILING DATE: 2003-03-05
; PRIOR APPLICATION NUMBER: GB 0218518.9
; PRIOR FILING DATE: 2002-09-08
; PRIOR APPLICATION NUMBER: US 60/366,601
; PRIOR FILING DATE: 2002-03-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 3951
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-379-981-8

Query Match 93.2%; Score 2249; DB 16; Length 3951;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2299; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 73 ATGACCATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTATCAGAATCAGCCT 132
Db 1 ATGACCATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTATCAGAATCAGCCT 60
Qy 133 GGCAGCTCCGAGGCGAGTCTCACCTGGAGACATGATGAGGTTCTGCCAGCTGGGGTCT 192
Db 61 GGCAGCTCCGAGGCGAGTCTCACCTGGAGACATGATGAGGTTCTGCCAGCTGGGGTCT 120
Qy 193 GTGCTTCATTGAAATGATGTGCAAAATCAGACACTTTCTTTAGGACAGTACCTGGTCT 252
Db 121 GTGCTTCATTGAAATGATGTGCAAAATCAGACACTTTCTTTAGGACAGTACCTGGTCT 180
Qy 253 GTAGTTATTTCAGTTTATCTGTATCTGTACCTGATAAATCAAAACCATCACCACAAAGATCAA 312
Db 181 GTAGTTATTTCAGTTTATCTGTATCTGTACCTGATAAATCAAAACCATCACCACAAAGATCAA 240
Qy 313 GCCCTAGTGTATGGTATCGCTCCTCCACAGAAAGTTCTTTTCCCATCTGAGAGATTTGT 372
Db 241 GCCCTAGTGTATGGTATCGCTCCTCCACAGAAAGTTCTTTTCCCATCTGAGAGATTTGT 300
Qy 373 CTTAAGTGGCAACAACACTCATAGTTCGAGCTGGCTCCAGAAATTTGGGCAATACCTGT 432
Db 301 CTTAAGTGGCAACAACACTCATAGTTCGAGCTGGCTCCAGAAATTTGGGCAATACCTGT 360
Qy 433 TTTGCAATGACGACTGCAAGTGTGTTAACTACACACACACCTTTGCGCAATACATGCTA 492
Db 361 TTTGCAATGACGACTGCAAGTGTGTTAACTACACACACACCTTTGCGCAATACATGCTA 420

QY 493 TCACATGAACACTCCAAAAACATGTCATCGAAGGCTTTTGTATGATGTCATCAATGCAA 552
Db 421 TCACATGAACACTCCAAAAACATGTCATCGAAGGCTTTTGTATGATGTCATCAATGCAA 480
QY 553 GCACATATTACCCAGGCACTCAGTAATCTCTGGGACGTTTATAACCAATGTTTGTATC 612
Db 481 GCACATATTACCCAGGCACTCAGTAATCTCTGGGACGTTTATAACCAATGTTTGTATC 540
QY 613 AATGAGATGGGGGTATAGCTAGGCACCTCGTTTGGAAACCAAGAGATGCCATGAA 672
Db 541 AATGAGATGGGGGTATAGCTAGGCACCTCGTTTGGAAACCAAGAGATGCCATGAA 600
QY 673 TTCTTCAATACACTGTTGATGCTATGACAGAAAGCATGCTTGAATGGCAGCAATAATTA 732
Db 601 TTCTTCAATACACTGTTGATGCTATGACAGAAAGCATGCTTGAATGGCAGCAATAATTA 660
QY 733 GACAGACACACCCAGGCACTCCTGTTTGTTCAGATATTGGAGGATACCTAAGATCT 792
Db 661 GACAGACACACCCAGGCACTCCTGTTTGTTCAGATATTGGAGGATACCTAAGATCT 720
QY 793 AGAGTCAAAATGTTTAAATGCAAGGCGTTTCAGATACCTTTTCATCCATATCTTGATATA 852
Db 721 AGAGTCAAAATGTTTAAATGCAAGGCGTTTCAGATACCTTTTCATCCATATCTTGATATA 780
QY 853 ACATTGGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCATTGGAGCAGTTTGTGAAGCG 912
Db 781 ACATTGGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCATTGGAGCAGTTTGTGAAGCG 840
QY 913 GAAACAGTTGATGGAGAAACTCGTACAGTGCAGCAAGTGTAAAGATGTTTCCAGCT 972
Db 841 GAAACAGTTGATGGAGAAACTCGTACAGTGCAGCAAGTGTAAAGATGTTTCCAGCT 900
QY 973 TCAAGAGAGTTCACTATCCATAGATCTCTTAATGTTTCTTACACTTCTCTGAAACGTTTT 1032
Db 901 TCAAGAGAGTTCACTATCCATAGATCTCTTAATGTTTCTTACACTTCTCTGAAACGTTTT 960
QY 1033 GCAATTTTACCGGTGGAAAAATGCTAAGGATGTGAATACCTCAGTATCTTGATATT 1092
Db 961 GCAATTTTACCGGTGGAAAAATGCTAAGGATGTGAATACCTCAGTATCTTGATATT 1020
QY 1093 CGGCATATATGTCACACCAACGAGAGCCAAATGTCAGTCTTGATCAGTCTG 1152
Db 1021 CGGCATATATGTCACACCAACGAGAGCCAAATGTCAGTCTTGATCAGTCTG 1080
QY 1153 GTCCACACTGTTTAAATGTCATGCTGCGCAATTAATCTGTCATATAAAGCTAGCAAT 1212
Db 1081 GTCCACACTGTTTAAATGTCATGCTGCGCAATTAATCTGTCATATAAAGCTAGCAAT 1140
QY 1213 GGCCTCTGGTATCAATGAATGACTCCATTTGATCTACAGTGTATATAGATCGGTACTC 1272
Db 1141 GGCCTCTGGTATCAATGAATGACTCCATTTGATCTACAGTGTATATAGATCGGTACTC 1200
QY 1273 AGCCAAACAGCCTATGCTCTTTATATCAGTCCCATGATGTGAAAAATGGAGGTGAA 1332
Db 1201 AGCCAAACAGCCTATGCTCTTTATATCAGTCCCATGATGTGAAAAATGGAGGTGAA 1260
QY 1333 CTTACTATCCACCCATAGCCCCCGGCGAGTCTCTCCCGCCCGCATCAGTCAGCGG 1392
Db 1261 CTTACTATCCACCCATAGCCCCCGGCGAGTCTCTCCCGCCCGCATCAGTCAGCGG 1320
QY 1393 GTTGTACCAACAAACAGGCTGCGCCAGGCTTTATCGSACACACAGCTTCCCTCTCAATG 1452
Db 1321 GTTGTACCAACAAACAGGCTGCGCCAGGCTTTATCGSACACACAGCTTCCCTCTCAATG 1380
QY 1453 ATAAGAATCCACCTCACTTAATGGGACTGGACCATTTGAAGACACGCCAAGCAGTTCC 1512
Db 1381 ATAAGAATCCACCTCACTTAATGGGACTGGACCATTTGAAGACACGCCAAGCAGTTCC 1440
QY 1513 ATGTGAGTCTTAACGGGAATTCAGATGTCAACAGGCTAGTCTGTTAATGCTTCACT 1572
Db 1441 ATGTGAGTCTTAACGGGAATTCAGATGTCAACAGGCTAGTCTGTTAATGCTTCACT 1500
QY 1573 TCTGTCAAAACTGGTCAGTTAATAGTCTCCTCAGTGTATCCAGAACATCCTAAGAACAA 1632

Db 1501 TCTGTCCAAAACCTGGTCAGTTAATAGGTCTCTCAGTGTATCCAGAACATCTCTAAGAACAA 1560
QY 1633 AAAATTACAATCAGTATTCAACAAGTTCGCTGTTGCGCAGTGTCACTCTCAACCTAAC 1692
Db 1561 AAAATTACAATCAGTATTCAACAAGTTCGCTGTTGCGCAGTGTCACTCTCAACCTAAC 1620
QY 1693 CTTTCATAGTAAATTTTGGAGAACCTTACCAAGCCCTTCCCTTCTTACCATPACCAAT 1752
Db 1621 CTTTCATAGTAAATTTTGGAGAACCTTACCAAGCCCTTCCCTTCTTACCATPACCAAT 1680
QY 1753 TCTGTCAGTACAGTCTACCTCGAAGCATCTACGATGTCACTTCTAGTAAAGTAAACAAA 1812
Db 1681 TCTGTCAGTACAGTCTACCTCGAAGCATCTACGATGTCACTTCTAGTAAAGTAAACAAA 1740
QY 1813 CCGATCCCCCGCAGTGAATCTCTCCAGCCGCTGATGATGATGATGATGATGATGATGATGAT 1872
Db 1741 CCGATCCCCCGCAGTGAATCTCTCCAGCCGCTGATGATGATGATGATGATGATGATGATGAT 1800
QY 1873 TCAGCGTCTGCTGCTGCTTATGGCGCCGAGTCTCTGAGGACTCTGACGAGGAGTCAAG 1932
Db 1801 TCAGCGTCTGCTGCTGCTTATGGCGCCGAGTCTCTGAGGACTCTGACGAGGAGTCAAG 1860
QY 1933 GGGCTGGGCAAGGAGAAATGGGATTTGATGATGATGATGATGATGATGATGATGATGAT 1992
Db 1861 GGGCTGGGCAAGGAGAAATGGGATTTGATGATGATGATGATGATGATGATGATGATGAT 1920
QY 1993 GCGAAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCCCTAAACGGTGT 2052
Db 1921 GCGAAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCCCTAAACGGTGT 1980
QY 2053 AATAGTGCAGACAGCAGACAGTGAACCGAAGAAACCGGCTAGCGCTGATGATGATGATGAT 2112
Db 1981 AATAGTGCAGACAGCAGACAGTGAACCGAAGAAACCGGCTAGCGCTGATGATGATGATGAT 2040
QY 2113 TSCCAAGGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2172
Db 2041 TGCCAAAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
QY 2173 GGAAGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2232
Db 2101 GGAAGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2160
QY 2233 AGGCTTAGCAACAACTGAAAGGCTCGACGGATGAATGATGATGATGATGATGATGATGAT 2292
Db 2161 AGGCTTAGCAACAACTGAAAGGCTCGACGGATGAATGATGATGATGATGATGATGATGAT 2220
QY 2293 GGCCTTCCCGAGGACCGGACGCGAGCCTCAGCTGCGAGCCCGCCGCGGAAATCCCTG 2352
Db 2221 GGCCTTCCCGAGGACCGGACGCGAGCCTCAGCTGCGAGCCCGCCGCGGAAATCCCTG 2280
QY 2353 GAGGAGCCAGATGCGGCCG 2372
Db 2281 GAGGAGCCAGATGCGGCCG 2300

RESULT 4

US-10-302-172-390
; Sequence 390, Application US/10302172
; Publication No. US20040053250A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20040053250A1 Arginine-rich Protein-like Nucleic Acids and
; FILE REFERENCE: 803_ICNCP
; CURRENT APPLICATION NUMBER: US/10/302,172
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/225,251
; PRIOR FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: PCT US02/05095
; PRIOR FILING DATE: 2002-03-05


```

; PRIOR APPLICATION NUMBER: US 09/799,451
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 950
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 390
; LENGTH: 4483
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (191)..(4000)
; US-10-302-172-390

Query Match      91.6%; Score 2210; DB 13; Length 4483;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2310; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 61 CAGAGTTGAACATGACCATAGTTGACCAAGCTTCTGAATCTTCAGACCCATCAGCCTAT 120
Db 179 CAGAGTTGAACATGACCATAGTTGACCAAGCTTCTGAATCTTCAGACCCATCAGCCTAT 238
QY 121 CAGAAATCAGCTGCGAGCTCCGAGGAGTCTCACCTGGAGACATGGATGAGGCTTCTGCC 180
Db 239 CAGAAATCAGCTGCGAGCTCCGAGGAGTCTCACCTGGAGACATGGATGAGGCTTCTGCC 298
QY 181 AGCTGGGGTCTGTGCTTCAATGAATGATGTGTCAAAATCACACATCTTCTTTAGGACCA 240
Db 299 AGCTGGGGTCTGTGCTTCAATGAATGATGTGTCAAAATCACACATCTTCTTTAGGACCA 358
QY 241 GTACCTGGTCTGTAGTTTATTCAGTTTCATCTGTACCTGATATAAATCAAAACCATCACCA 300
Db 359 GTACCTGGTCTGTAGTTTATTCAGTTTCATCTGTACCTGATATAAATCAAAACCATCACCA 418
QY 301 CAAAGAGATCAAGCCCTAGGTGATGGCATCGCTCCTCCACAGAAAGTTCTTTTCCCATCT 360
Db 419 CAAAGAGATCAAGCCCTAGGTGATGGCATCGCTCCTCCACAGAAAGTTCTTTTCCCATCT 478
QY 361 GAGAAGATTTTGTCTTAAGTGGCAACAAATCATAGATGTGGAGCTGGGCTCCAGAAATTG 420
Db 479 GAGAAGATTTTGTCTTAAGTGGCAACAAATCATAGATGTGGAGCTGGGCTCCAGAAATTG 538
QY 421 GGCAATACCTGTTTGGCAATGGAGCATGAGCTGAGTGTAACTACACACACCTCTTGCC 480
Db 539 GGCAATACCTGTTTGGCAATGGAGCATGAGCTGAGTGTAACTACACACACCTCTTGCC 598
QY 481 AATTACATGCTATCACATGAACACTCCAAACATGTCTATGAGAGGCTTTTGTATGATG 540
Db 599 AATTACATGCTATCACATGAACACTCCAAACATGTCTATGAGAGGCTTTTGTATGATG 658
QY 541 TGTAATGCAAGCACATATTACCCAGGACATCAGTAATCTCTGGGAGCTTATTAAACCA 600
Db 659 TGTAATGCAAGCACATATTACCCAGGACATCAGTAATCTCTGGGAGCTTATTAAACCA 718
QY 601 ATGTTTGTATCAATGAGATGGGGTATAGTGGGACCTCCGTTTGGAAACCAAGAA 660
Db 719 ATGTTTGTATCAATGAGATGGGGTATAGTGGGACCTCCGTTTGGAAACCAAGAA 778
QY 661 GATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGCTAGAAAGCATGCTTGAATGCG 720
Db 779 GATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGCTAGAAAGCATGCTTGAATGCG 838
QY 721 AGCAATAAATTAGACAGACACACCCAGGACCATCTCTGTTTGTTCAGATATTTGGAGGA 780
Db 839 AGCAATAAATTAGACAGACACACCCAGGACCATCTCTGTTTGTTCAGATATTTGGAGGA 898
QY 781 TACCTAAGATCTAGAGTCAAAATGTTTAAATTCAGAGGGGTTTTCAGATATTTGATCCA 840
Db 899 TACCTAAGATCTAGAGTCAAAATGTTTAAATTCAGAGGGGTTTTCAGATATTTGATCCA 958
QY 841 TATCTTGTATATAACATTTGGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCATTTGGAGCAG 900
Db 959 TATCTTGTATATAACATTTGGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCATTTGGAGCAG 1018

```

```

QY 901 TTTTGTGAAGCCGGAACAGCTTGATCGAGAAAACTCGTACAGTGCAGCAAGTGTAAAGAG 960
Db 1019 TTTTGTGAAGCCGGAACAGCTTGATCGAGAAAACTCGTACAGTGCAGCAAGTGTAAAGAG 1078
QY 961 ATGGTTCAGCTTCAAGAGGTTCACTATCCATAGATCCTCTAAATGTTCTTACACTTTCT 1020
Db 1079 ATGGTTCAGCTTCAAGAGGTTCACTATCCATAGATCCTCTAAATGTTCTTACACTTTCT 1138
QY 1021 CTGAAACGTTTTTCAAAATTTTACCGGTGGAATAAATTTGCTTAAGGATGTGAAATCCCTGAG 1080
Db 1139 CTGAAACGTTTTTCAAAATTTTACCGGTGGAATAAATTTGCTTAAGGATGTGAAATCCCTGAG 1198
QY 1081 TATCTTGATATTTCGGCCATATATGCTCAACCAACGAGAGCCATTTCTACGTTCTTG 1140
Db 1199 TATCTTGATATTTCGGCCATATATGCTCAACCAACGAGAGCCATTTCTACGTTCTTG 1258
QY 1141 TATGAGTCTGTGTCACACTGCTTTTAAATGCAATGCTTGGCCATTTACTTCTGCTACATA 1200
Db 1259 TATGAGTCTGTGTCACACTGCTTTTAAATGCAATGCTTGGCCATTTACTTCTGCTACATA 1318
QY 1201 AAGACTAGCAATGGCCTCTGGTATCAAATGAATGACTTCCATTTGATCTACAGTGATATT 1260
Db 1319 AAGACTAGCAATGGCCTCTGGTATCAAATGAATGACTTCCATTTGATCTACAGTGATATT 1378
QY 1261 AGATCGGTACTCAGCAACAAAGCCTATGCTCTTTTATATCATAGTCCCATGATGTGAAA 1320
Db 1379 AGATCGGTACTCAGCAACAAAGCCTATGCTCTTTTATATCATAGTCCCATGATGTGAAA 1438
QY 1321 AATGGAGGTGAACCTTACTCATCCACATAGCCCGGCCAGTCTCTCCCGCCCGCTC 1380
Db 1439 AATGGAGGTGAACCTTACTCATCCACATAGCCCGGCCAGTCTCTCCCGCCCGCTC 1498
QY 1381 ATCAGTCAGCGGGTGTGTCACCAACAAAGCCTGCGCCAGGCTTTATCGGACCAACAGCTT 1440
Db 1499 ATCAGTCAGCGGGTGTGTCACCAACAAAGCCTGCGCCAGGCTTTATCGGACCAACAGCTT 1558
QY 1441 CCCTCTCACATGATPAAAGAAATCCACCTCATTAAATGGGACTGACCATTTGAAGAGACAG 1500
Db 1559 CCCTCTCACATGATPAAAGAAATCCACCTCATTAAATGGGACTGACCATTTGAAGAGACAG 1618
QY 1501 CCAAGCAGTTCCTCAGTGCAGTCTTAACGGGAATTCAGTGTCAACAGGGCTAGTCTGTT 1560
Db 1619 CCAAGCAGTTCCTCAGTGCAGTCTTAACGGGAATTCAGTGTCAACAGGGCTAGTCTGTT 1678
QY 1561 AATGCTTCAGCTTCTGTCCTCAAACTGCTCAGTTAATAGTCTCCTCAGTGATCCGAGACAT 1620
Db 1679 AATGCTTCAGCTTCTGTCCTCAAACTGCTCAGTTAATAGTCTCCTCAGTGATCCGAGACAT 1738
QY 1621 CCTAAGAAACAAANAATTAACATCAGTATTCAACACAGATTGCTGTTGCGCAGTGTGAG 1680
Db 1739 CCTAAGAAACAAANAATTAACATCAGTATTCAACACAGATTGCTGTTGCGCAGTGTGAG 1798
QY 1681 TCTCAACCTTAACCTTCATAGTAATTTCTTTGGAGAACCTTACCAAGCCCGTTCCCTCTTCT 1740
Db 1799 TCTCAACCTTAACCTTCATAGTAATTTCTTTGGAGAACCTTACCAAGCCCGTTCCCTCTTCT 1858
QY 1741 ACCATTACCAATTCCTGAGTACAGTCTACCTCGAAGCATCTAAGATGTGATTTCTAGT 1800
Db 1859 ACCATTACCAATTCCTGAGTACAGTCTACCTCGAAGCATCTAAGATGTGATTTCTAGT 1918
QY 1801 AAAGTAAACAAACCATCCCGGAGTGAATCTCTCCAGCCGCTGATGAATGGCAAA 1860
Db 1919 AAAGTAAACAAACCATCCCGGAGTGAATCTCTCCAGCCGCTGATGAATGGCAAA 1978
QY 1861 TCCAGAGTGAACCTCCAGGCTGCTGCTGCTATGGCGCCAGTCTCTGAGGACTCTGAC 1920
Db 1979 TCCAGAGTGAACCTCCAGGCTGCTGCTGCTATGGCGCCAGTCTCTGAGGACTCTGAC 2038
QY 1921 GAGGAGTCAAAAGGGCTGGGCAAGGAGATGGGATGGAATGGAATGTGAGTCTCCACTCT 1980
Db 2039 GAGGAGTCAAAAGGGCTGGGCAAGGAGATGGGATGGAATGGAATGTGAGTCTCCACTCT 2098
QY 1981 CCGGCGCAAGATGCGAAGATGAGGAGGCCACTCCCGACGAGCTTCAAGAACCCATGACC 2040

```

```
Db 2099 CCCGCCAAGATCCGAAGATGAGGAGCCACTCCGACAGCTTCAAGAACCCTACCC 2158
QY 2041 CTAACCGTCTAATAGTCAGACAGCGACAGTACCCGAAAGAAAGCGCCTAGCGCCT 2100
Db 2159 CTAACCGTCTAATAGTCAGACAGCGACAGTACCCGAAAGAAAGCGCCTAGCGCCT 2218
QY 2101 GATGGTCCAGCTGCCAAGGCCAGCCTGCCCTGCCTTCTCTCCAGAGCAAAATC 2160
Db 2219 GATGGTCCAGCTGCCAAGGCCAGCCTGCCCTGCCTTCTCTCCAGAGCAAAATC 2278
QY 2161 AACGGTCTTCTGAAAGTTGATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2220
Db 2279 AACGGTCTTCTGAAAGTTGATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2338
QY 2221 TTAGAGACCTTCAGGCTTACCAACAATGAAAGGCTCGACGATGAATGATGACCT 2280
Db 2339 TTAGAGACCTTCAGGCTTACCAACAATGAAAGGCTCGACGATGAATGATGACCT 2398
QY 2281 GGAGCAGAGAGGGGCCCTCCGAGGACCGCGAGCCTCAGCTGGCAGCCCGCC 2340
Db 2399 GGAGCAGAGAGGGGCCCTCCGAGGACCGCGAGCCTCAGCTGGCAGCCCGCC 2458
QY 2341 GCGGATCCCTGGAGGAGCGACGATGCGGCGC 2372
Db 2459 GCGGATCCCTGGAGGAGCGACGATGCGGCGC 2490
```

RESULT 5

```
US-09-864-761-31553/c
; Sequence 31553, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
```

```
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31553
; LENGTH: 333
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004895.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.72
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.66
; OTHER INFORMATION: SWISSPROT HIT: P51612, EVALUE 4.90e+00
; OTHER INFORMATION: NT HIT: G17305322, EVALUE 2.10e-01
; OTHER INFORMATION: EST_HUMAN HIT: AW298129.1, EVALUE 0.00e+00
US-09-864-761-31553
```

```
Query Match 13.8%; Score 333; DB 9; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.1e-168;
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1843 CCCGTGATGAATGGCAAAATCCAAAGTGAACCTCCAGGCTGCTGTCCTATGCGCCGAG 1902
Db 333 CCCGTGATGAATGGCAAAATCCAAAGTGAACCTCCAGGCTGCTGTCCTATGCGCCGAG 274
QY 1903 TCCTCTGAGGACTCTGACGAGGAGTCAAAGGGGCTGGGCAAGAGAAATGGGATTGGTAGC 1962
Db 273 TCCTCTGAGGACTCTGACGAGGAGTCAAAGGGGCTGGGCAAGAGAAATGGGATTGGTAGC 214
QY 1963 ATTGTGAGTCCCACTCTCCCGGCCAAGATGCCAAGATGAGGAGGCCACTCCGACCGAG 2022
Db 213 ATTGTGAGTCCCACTCTCCCGGCCAAGATGCCAAGATGAGGAGGCCACTCCGACCGAG 154
QY 2023 CTTCAAGAACCCATGACCTTAAACGGTCTTAATAGTCAGACAGCAGACAGTACCCGAAA 2082
Db 153 CTTCAAGAACCCATGACCTTAAACGGTCTTAATAGTCAGACAGCAGACAGTACCCGAAA 94
QY 2083 GAAAACGGCCTAGCGCCTGATGGTGCCAGCTGCCAAGGCCAGCCCTGCGCTGCACCTCAGAA 2142
Db 93 GAAAACGGCCTAGCGCCTGATGGTGCCAGCTGCCAAGGCCAGCCCTGCGCTGCACCTCAGAA 34
QY 2143 AATCCCTTTGCTAAGGCAACGGTCTTCCCTGGA 2175
Db 33 AATCCCTTTGCTAAGGCAACGGTCTTCCCTGGA 1
```

RESULT 6

```
US-09-864-761-15021/c
; Sequence 15021, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
```

```

; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 15021
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004895.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.72
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.66
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.74
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.63
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 0.88
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.77
; US-09-864-761-14092

Query Match 13.0%; Score 314; DB 9; Length 498;
Best Local Similarity 100.0%; Pred. No. 2.1e-158;
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1865 AGCTGAACCTCAGCGTGTGTGCTTATGGCGGAGTCTCTTGAGGACTCTGACGAGG 1924
Db 498 AGCTGAACCTCAGCGTGTGTGCTTATGGCGGAGTCTCTTGAGGACTCTGACGAGG 439

QY 1925 AGTCAAGGGCTGGCAGGAGATGGATTGTGATGCTCCACTCTCCCG 1984
Db 438 AGTCAAGGGCTGGCAGGAGATGGATTGTGATGCTCCACTCTCCCG 379

QY 1985 GCCAAGATGCCAAGATGAGAGGCGCACTCCGACGAGCTTCAAGAACCCATGACCCCTAA 2044
Db 378 GCCAAGATGCCAAGATGAGAGGCGCACTCCGACGAGCTTCAAGAACCCATGACCCCTAA 319

QY 2045 ACGGTGCTAATAGTCAGACGAGCAGTGCACCCGAAAGAAACGGCTAGGCGCTGATG 2104
Db 318 ACGGTGCTAATAGTCAGACGAGCAGTGCACCCGAAAGAAACGGCTAGGCGCTGATG 259

QY 2105 GTGCCAGCTGCAAGCGCAGCTGCGCTGCACTCAGAAAATCCCTTTGCTAAGGCAAAAG 2164
Db 258 GTGCCAGCTGCAAGCGCAGCTGCGCTGCACTCAGAAAATCCCTTTGCTAAGGCAAAAG 199

QY 2165 GTCTTCCTGGAAG 2178
Db 198 GTCTTCCTGGAAG 185
```

```

RESULT 7
US-09-864-761-14092/c
; Sequence 14092, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 14092
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004895.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.74
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.63
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 0.88
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.77
; US-09-864-761-14092

Query Match 8.5%; Score 205; DB 9; Length 489;
Best Local Similarity 100.0%; Pred. No. 1.8e-99;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 312 AGCCCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCATCTCAGAGATTG 371
Db |||||
449 AGCCCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCATCTCAGAGATTG 390
QY 372 TCTTAAGTGGCAACAACCTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGGCAATACCTG 431
Db |||||
389 TCTTAAGTGGCAACAACCTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGGCAATACCTG 330
QY 432 TTTTGCAATGACGACCTGAGTGTAACTACACACCACTCTTCCCAATTACATGCT 491
Db |||||
329 TTTTGCAATGACGACCTGAGTGTAACTACACACCACTCTTCCCAATTACATGCT 270
QY 492 ATCATGTAACACTCCAAAACATGT 516
Db |||||
269 ATCATGTAACACTCCAAAACATGT 245

RESULT 8
US-09-864-761-30650/c
; Sequence 30650, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 30650
; LENGTH: 198
; TYPE: DNA

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004895.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.74
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.63
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.88
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.77
; OTHER INFORMATION: NT HIT: AB040886.1, EVALUE 2.50e-02
; OTHER INFORMATION: EST HUMAN HIT: AW504799.1, EVALUE 1.00e-108
; OTHER INFORMATION: SWISSPROT HIT: Q61068, EVALUE 2.00e-13
US-09-864-761-30650

Query Match 8.2%; Score 198; DB 9; Length 198;
Best Local Similarity 100.0%; Pred. No. 1.1e-95;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 314 CCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCATCTCAGAGATTGTC 373
Db 198 CCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCATCTCAGAGATTGTC 139
QY 374 TTAAGTGGCAACAACCTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGCAATACCTGTT 433
Db 138 TTAAGTGGCAACAACCTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGCAATACCTGTT 79
QY 434 TTGCCAATGCACACTGCAGTGTAACTACACACCACTCTTCCCAATTACATGCTAT 493
Db 78 TTGCCAATGCACACTGCAGTGTAACTACACACCACTCTTCCCAATTACATGCTAT 19
QY 494 CACATGAACACTCCAAA 511
Db 18 CACATGAACACTCCAAA 1

RESULT 9
US-10-027-632-128282/c
; Sequence 128282, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 128282
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-128282

Query Match 6.5%; Score 157; DB 13; Length 504;
Best Local Similarity 100.0%; Pred. No. 1.6e-73;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1302	CAGGTCCTCATGATGTGAAAATCGAGGTGAATTACTCATCCACCATAGCCCCGGCCA	1361
Db	485	CAGGTCCTCATGATGTGAAAATCGAGGTGAATTACTCATCCACCATAGCCCCGGCCA	426
QY	1362	GTCTCTTCCCGCCCGGTCAATAGTCAGCGGGTTTCACCAACAAACAGCTCGGCCAGG	1421
Db	425	GTCTCTTCCCGCCCGGTCAATAGTCAGCGGGTTTCACCAACAAACAGCTCGGCCAGG	366
QY	1422	CTTTATCGGACCAAGCTTCCCTCTCATATGATAAAG	1458
Db	365	CTTTATCGGACCAAGCTTCCCTCTCATATGATAAAG	329

```

RESULT 10
US10-027-632-128282/c
; Requirement 128282; Application US/10027632
; Registration No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome.
; TITLE OF INVENTION: Polymorphisms in the Human Genome.

```

```

, CURRENT APPLICATION NUMBER: US/10/027,632
, CURRENT FILING DATE: 2002-04-30
, PRIOR APPLICATION NUMBER: US 60/218,006
, PRIOR FILING DATE: 2000-07-12
, PRIOR APPLICATION NUMBER: US 60/198,676
, PRIOR FILING DATE: 2000-04-20
, PRIOR APPLICATION NUMBER: US 60/193,483
, PRIOR FILING DATE: 2000-03-29
, PRIOR APPLICATION NUMBER: US 60/185,218
, PRIOR FILING DATE: 2000-02-24
, PRIOR APPLICATION NUMBER: US 60/167,363
, PRIOR FILING DATE: 1999-11-23
, PRIOR APPLICATION NUMBER: US 60/156,358
, PRIOR FILING DATE: 1999-09-28
, PRIOR APPLICATION NUMBER: US 60/146,002
, PRIOR FILING DATE: 1999-08-09
, NUMBER OF SEQ ID NOS: 325720
, SOFTWARE: FastSEQ for Windows Version 4.0
, SEQ ID NO 128282

```

	Query Match	6.5%;	Score 157;	DB 16;	Length 504;
	Best Local Similarity	100.0%;	Pred. No. 1.6e-73;		
	Matches 157;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
QY	1302	CAGGTCCCATGATGTGAAAAATGGAGGTTGA	CTTACTCATCCACCCATAGCCCCGGCCA	1366	
DB	485	CAGGTCCCATGATGTGAAAAATGGAGGTTGA	CTTACTCATCCACCCATAGCCCCGGCCA	426	
QY	1362	GTTCCTCTCCCGCCCGTTCATCAGTCAGCGGGT	TGTTCACCAACAAACAGGTCGCCAGG	1422	
DB	425	GTTCCTCTCCCGCCCGTTCATCAGTCAGCGGGT	TGTTCACCAACAAACAGGTCGCCAGG	366	
QY	1422	CTTTATCGGACCACAGCTTCCCTCTCACATGATAA	AG	1458	
DB	365	CTTTATCGGACCACAGCTTCCCTCTCACATGATAA	AG	329	

```

RESULT 11
US-10-027-632-277717
; Sequence 277717, Application US/10027632
; Publication NO. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129

```

```

, CURRENT APPLICATION NUMBER: US/10/027,632
,
, CURRENT FILING DATE: 2002-04-30
, PRIOR APPLICATION NUMBER: US 60/218,006
, PRIOR FILING DATE: 2000-07-12
, PRIOR APPLICATION NUMBER: US 60/198,676
, PRIOR FILING DATE: 2000-04-20
, PRIOR APPLICATION NUMBER: US 60/193,483
, PRIOR FILING DATE: 2000-03-29
, PRIOR APPLICATION NUMBER: US 60/185,218
, PRIOR FILING DATE: 2000-02-24
, PRIOR APPLICATION NUMBER: US 60/167,363
, PRIOR FILING DATE: 1999-11-23
, PRIOR APPLICATION NUMBER: US 60/156,358
, PRIOR FILING DATE: 1999-09-28
, PRIOR APPLICATION NUMBER: US 60/146,002
, PRIOR FILING DATE: 1999-08-09
, NUMBER OF SEQ ID NOS: 325720
, SOFTWARE: fastseq for Windows Version 4.0
, SEQ ID NO 277717
, LENGTH: 631
, TYPE: DNA
, ORGANISM: Human
, US-10-027-632-277717

```

	Query Match	4.0%;	Score 97;	DB 13;	Length 631;
	Best Local Similarity	100.0%;	Pred. NO. 4.4e-41;		
	Matches 97;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	2177	AGTTGATGCGCTGCTCCTTTGCTGCTCTCCGAGAAGCAAAATCTTAGAGACCTTCAGGC	2236		
DB	175	AGTTGATGCGCTGCTCCTTTGCTGCTCTCCGAGAAGCAAAATCTTAGAGACCTTCAGGC	234		
QY	2237	TTAGCAACAAACTGAAAGGCTCGACGGATGAAATGAG	2273		
DB	235	TTAGCAACAAACTGAAAGGCTCGACGGATGAAATGAG	271		

```

RESULT 12
US-10-027-632-277717
; Sequence 277717, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108927.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-01-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-03-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 277717
; LENGTH: 631
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-277717

```

Query Match	4.0%	Score 97;	DB 16;	Length 631;
Best Local Similarity	100.0%	Pred. No. 4.4e-41;		
Matches 97; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

```

RESULT 16
US-10-027-632-319601/c
; Sequence 319601, Application US/10027632
; Publication No. US5002019837A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Ma
; TITLE OF INVENTION: Polymorphisms in the
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632

```

Query Match 1.6%; Score 39; DB 16; Length 592;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 758 TTGTTGTCAGATATTTGGAGGATACCTAAGATCTAGAG 796
DB 460 TTGTTGTCAGATATTTGGAGGATACCTAAGATCTAGAG 422

RESULT 18
US-10-258-662-51/c
; Sequence 51, Application US/10258662
; Publication No. US20040082029A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: LAL, Preeti
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: BAUGHN, Mariat R.
; APPLICANT: YAO, Monique G.
; APPLICANT: BURFORD, Neil
; APPLICANT: BATRA, Sajeev
; APPLICANT: POLICKY, Jennifer J.
; TITLE OF INVENTION: RNA METABOLISM PROTEINS
; FILE REFERENCE: PF-0771 USN
; CURRENT APPLICATION NUMBER: US/10/258,662
; CURRENT FILING DATE: 2002-10-25
; PRIOR APPLICATION NUMBER: US 60/201,875
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/200,184
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/202,090
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/210,232
; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: US 60/220,553
; PRIOR FILING DATE: 2000-07-25
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PERL Program
; SEQ ID NO 51
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 1647264CB1
US-10-258-662-51

Query Match 1.2%; Score 30; DB 17; Length 637;
Best Local Similarity 100.0%; Pred. No. 7.4e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAATTTAGCTTGA 2412
DB 37 CCCTTTAGTGAGGGTTAATTTAGCTTGA 8

RESULT 19
US-10-149-819-39
; Sequence 39, Application US/10149819
; Publication No. US20030044913A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: TANG, Y. Tom
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BAUGHN, Marian R.
; APPLICANT: LU, Dyung Aina M.

Wed Aug 18 13:52:59 2004

US-10-049-745-31.oli30.rnpb

Query Match 1.6%; Score 39; DB 13; Length 592;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 758 TTGTTGTCAGATATTTGGAGGATACCTAAGATCTAGAG 796
DB 460 TTGTTGTCAGATATTTGGAGGATACCTAAGATCTAGAG 422

RESULT 17
US-10-027-632-319601/c
; Sequence 319601, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 319601
; LENGTH: 592
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(592)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-319601

; APPLICANT: SHAH, Purvi
; APPLICANT: LAL, Preeti
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: BURFORD, Neil
; TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
; FILE REFERENCE: PF-0760 PCT
; CURRENT APPLICATION NUMBER: US/10/149,819
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
; PRIOR FILING DATE: 1999-12-10; 1999-12-16
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 39
; LENGTH: 733
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030044913A1 3683905CB1
US-10-149-819-39

Query Match 1.2%; Score 30; DB 15; Length 733;
Best Local Similarity 100.0%; Pred. No. 7.5e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
|||||
Db 696 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 725

RESULT 20

US-10-313-542-40/c
; Sequence 40, Application US/10313542
; Publication No. US20030120057A1
; GENERAL INFORMATION:
; APPLICANT: Roopa, Reddy
; APPLICANT: Guegler, Karl, J.
; APPLICANT: Au-Young, Janice
; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED P
; FILE REFERENCE: PA-0013 US
; CURRENT APPLICATION NUMBER: US/10/313,542
; CURRENT FILING DATE: 2002-12-05
; PRIOR APPLICATION NUMBER: US/09/495,050
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/118,318
; PRIOR FILING DATE: 1999-02-01
; NUMBER OF SEQ ID NOS: 305
; SOFTWARE: PERL Program
; SEQ ID NO 40
; LENGTH: 847
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030120057A1 797777CT1
US-10-313-542-40

Query Match 1.2%; Score 30; DB 15; Length 847;
Best Local Similarity 100.0%; Pred. No. 7.5e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
|||||
Db 37 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 8

RESULT 21

US-09-349-015-17
; Sequence 17, Application US/09349015A
; Patent No. US20020015950A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen Anne
; APPLICANT: Volkmut, Wayne

; APPLICANT: Walker, Michael
; TITLE OF INVENTION: ATHEROSCLEROSIS-ASSOCIATED GENES
; FILE REFERENCE: PB-0013 US
; CURRENT APPLICATION NUMBER: US/09/349,015A
; CURRENT FILING DATE: 1999-07-07
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 972
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-349-015-17

Query Match 1.2%; Score 30; DB 9; Length 972;
Best Local Similarity 100.0%; Pred. No. 7.5e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
|||||
Db 936 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 965

RESULT 22

US-10-219-664-13
; Sequence 13, Application US/10219664
; Publication No. US20030129176A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen Anne
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael G.
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: ATHEROSCLEROSIS-ASSOCIATED GENES
; FILE REFERENCE: PB-0013-1 CIP
; CURRENT APPLICATION NUMBER: US/10/219,664
; CURRENT FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: 09/349,015
; PRIOR FILING DATE: 1999-07-07
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PERL Program
; SEQ ID NO 13
; LENGTH: 972
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
US-10-219-664-13

Query Match 1.2%; Score 30; DB 15; Length 972;
Best Local Similarity 100.0%; Pred. No. 7.5e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
|||||
Db 936 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 965

RESULT 23

US-09-880-192-16
; Sequence 16, Application US/09880192
; Patent No. US20020077470A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION

; FILE REFERENCE: PB-0009-1 CIP
; CURRENT APPLICATION NUMBER: US/09/880,192
; CURRENT FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 16
; LENGTH: 1112

US-09-981-353-187		US-10-049-745-31.oli30.rnpb	
Query Match 1.2%; Score 30; DB 9; Length 1273; Best Local Similarity 100.0%; Pred. No. 7.6e-05; Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412 	QY	2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412
DB	1238 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 1267 	DB	1238 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 1267
RESULT 26		RESULT 27	
US-09-919-039-191 ; Sequence 191, Application US/09919039 ; Publication No. US20030108871A1 ; GENERAL INFORMATION: ; APPLICANT: Kaser, Matthew R. ; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES ; FILE REFERENCE: PA-0035 US ; CURRENT APPLICATION NUMBER: US/09/919,039 ; CURRENT FILING DATE: 2002-09-09 ; PRIOR APPLICATION NUMBER: 60/222,113 ; PRIOR FILING DATE: 2000-07-28 ; NUMBER OF SEQ ID NOS: 401 ; SOFTWARE: PERL Program ; SEQ ID NO 191 ; LENGTH: 1273 ; TYPE: DNA ; ORGANISM: Homo sapiens ; FEATURE: ; NAME/KEY: misc feature ; OTHER INFORMATION: Incyte ID No. US20030108871A1 522433CB1		US-10-133-013-59 ; Sequence 59, Application US/10133013 ; Publication No. US20030166903A1 ; GENERAL INFORMATION: ; APPLICANT: Astromoff, Anna ; APPLICANT: Cocks, Benjamin G. ; APPLICANT: Bandman, Olga ; TITLE OF INVENTION: GENES ASSOCIATED WITH VASCULAR DISEASE ; FILE REFERENCE: PA-0049 US ; CURRENT APPLICATION NUMBER: US/10/133,013 ; CURRENT FILING DATE: 2003-04-25 ; PRIOR APPLICATION NUMBER: 60/287,067 ; PRIOR FILING DATE: 2001-04-27 ; NUMBER OF SEQ ID NOS: 271 ; SOFTWARE: PERL Program ; SEQ ID NO 59 ; LENGTH: 1273 ; TYPE: DNA ; ORGANISM: Homo sapiens ; FEATURE: ; NAME/KEY: misc feature ; OTHER INFORMATION: Incyte ID No. US20030166903A1 522433CB1	
US-09-919-039-191		US-10-133-013-59	
Query Match 1.2%; Score 30; DB 10; Length 1273; Best Local Similarity 100.0%; Pred. No. 7.6e-05; Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		Query Match 1.2%; Score 30; DB 15; Length 1273; Best Local Similarity 100.0%; Pred. No. 7.6e-05; Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412 	QY	2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412
DB	1238 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 1267 	DB	1238 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 1267
RESULT 25		RESULT 24	
US-09-981-353-187 ; Sequence 187, Application US/09981353 ; Patent No. US20020160382A1 ; GENERAL INFORMATION: ; APPLICANT: Lasek, Amy W. ; APPLICANT: Jones, David A. ; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER ; FILE REFERENCE: PA-0038 US ; CURRENT APPLICATION NUMBER: US/09/981,353 ; CURRENT FILING DATE: 2001-10-11 ; NUMBER OF SEQ ID NOS: 194 ; SOFTWARE: PERL Program ; SEQ ID NO 187 ; LENGTH: 1273 ; TYPE: DNA ; ORGANISM: Homo sapiens ; FEATURE: ; NAME/KEY: misc feature ; OTHER INFORMATION: Incyte ID No. US20020160382A1 522433CB1		US-10-427-348-16 ; Sequence 16, Application US/10427348 ; Publication No. US20030175795A1 ; GENERAL INFORMATION: ; APPLICANT: Walker, Michael G. ; APPLICANT: Volkmut, Wayne ; APPLICANT: Klingler, Tod M. ; APPLICANT: Azimzai, Valda ; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION ; FILE REFERENCE: PB-0009-2 CON ; CURRENT APPLICATION NUMBER: US/10/427,348 ; CURRENT FILING DATE: 2003-04-29 ; PRIOR APPLICATION NUMBER: US 09/880,192 ; PRIOR FILING DATE: 2001-06-12 ; PRIOR APPLICATION NUMBER: US 09/299,708 ; PRIOR FILING DATE: 1999-04-26 ; NUMBER OF SEQ ID NOS: 62 ; SOFTWARE: PERL Program ; SEQ ID NO 16 ; LENGTH: 1112 ; TYPE: DNA ; ORGANISM: Homo sapiens ; FEATURE: ; NAME/KEY: misc feature ; OTHER INFORMATION: Incyte ID No. US20030175795A1 2951269CT1	
US-09-981-353-187		US-10-427-348-16	
Query Match 1.2%; Score 30; DB 9; Length 1112; Best Local Similarity 100.0%; Pred. No. 7.6e-05; Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		Query Match 1.2%; Score 30; DB 15; Length 1112; Best Local Similarity 100.0%; Pred. No. 7.6e-05; Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412 	QY	2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412
DB	1075 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 1104 	DB	1075 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 1104

```
Db 1238 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 1267
|||||
RESULT 28
US-09-349-015-2
; Sequence 2, Application US/09349015A
; Patent No. US20020015950A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen Anne
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael
; TITLE OF INVENTION: ATHEROSCLEROSIS-ASSOCIATED GENES
; FILE REFERENCE: PB-0013 US
; CURRENT APPLICATION NUMBER: US/09/349,015A
; CURRENT FILING DATE: 1999-07-07
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1702
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-349-015-2
Query Match 1.2%; Score 30; DB 9; Length 1702;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 2412
Db 1665 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 1694
|||||
RESULT 29
US-10-219-664-2
; Sequence 2, Application US/10219664
; Publication No. US20030129176A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen Anne
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael G.
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: ATHEROSCLEROSIS-ASSOCIATED GENES
; FILE REFERENCE: PB-0013-1 CIP
; CURRENT APPLICATION NUMBER: US/10/219,664
; CURRENT FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: 09/349,015
; PRIOR FILING DATE: 1999-07-07
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 1702
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
US-10-219-664-2
Query Match 1.2%; Score 30; DB 15; Length 1702;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 2412
Db 1665 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 1694
|||||
RESULT 30
US-09-880-192-32/c
; Sequence 32, Application US/09880192
; Patent No. US20020077470A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Jones, Karen Anne
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-1 CIP
; CURRENT APPLICATION NUMBER: US/09/880,192
; CURRENT FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 32
; LENGTH: 1833
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020077470A1 3795510CT1
US-10-427-348-32/c
; Sequence 32, Application US/10427348
; Publication No. US20030175795A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-2 CON
; CURRENT APPLICATION NUMBER: US/10/427,348
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 09/880,192
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: US 09/299,708
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 32
; LENGTH: 1833
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030175795A1 3795510CT1
US-10-427-348-32
Query Match 1.2%; Score 30; DB 15; Length 1833;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 2412
Db 37 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 8
|||||
RESULT 31
US-10-427-348-32/c
; Sequence 32, Application US/10427348
; Publication No. US20030175795A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-2 CON
; CURRENT APPLICATION NUMBER: US/10/427,348
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 09/880,192
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: US 09/299,708
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 32
; LENGTH: 1833
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030175795A1 3795510CT1
US-10-427-348-32
Query Match 1.2%; Score 30; DB 15; Length 1833;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 2412
Db 37 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 8
|||||
RESULT 32
US-09-971-392-216/c
; Sequence 216, Application US/09971392
; Publication No. US20030134283A1
; GENERAL INFORMATION:
; APPLICANT: Peterson, David P.
; APPLICANT: Pearson, Cecelia I.
; APPLICANT: Cocks, Benjamin G.
; TITLE OF INVENTION: GENES REGULATED IN DENDRITIC CELL DIFFERENTIATION
```

```
; FILE REFERENCE: PA-0029 US
; CURRENT APPLICATION NUMBER: US/09/971,392
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: 60/237,652
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PERL Program
; SEQ ID NO 216
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Template ID: 1095059.19
US-09-971-392-216

Query Match      1.2%; Score 30; DB 10; Length 1926;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
          |||
Db      36 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 7

RESULT 33
US-09-818-143-20/c
; Sequence 20, Application US/09818143
; Patent No. US20020019000A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; TITLE OF INVENTION: POLYNUCLEOTIDES COEXPRESSED WITH MATRIX-REMODELING GENES
; FILE REFERENCE: PB-0004 CIP
; CURRENT APPLICATION NUMBER: US/09/818,143
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PERL Program
; SEQ ID NO 20
; LENGTH: 1987
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 3948614CB1
US-09-818-143-20

Query Match      1.2%; Score 30; DB 9; Length 1987;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
          |||
Db      39 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 10

RESULT 34
US-10-198-846-10830/c
; Sequence 10830, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
```

```
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10830
; LENGTH: 2040
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032,
; LOCATION: 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-10830

Query Match      1.2%; Score 30; DB 15; Length 2040;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
          |||
Db      39 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 10

Search completed: August 17, 2004, 12:08:22
Job time : 3508 secs
```